Sevgi Dolu Zarafet Makineleri

Dario Amodei

Çev. Abdulhalim Durma

Dario Amodei, Anthropic'in kurucusu ve CEO'sudur. Şirket, güvenli, kontrol edilebilir ve yorumlanabilir yapay zeka sistemleri geliştirmeyi hedefler. OpenAI'den 2020'de ayrıldıktan sonra, kız kardeşi Daniela Amodei ve diğer ekiplerle birlikte Anthropic'ı kurdu. Amodei, AI güvenlik ve etik konularında öncü bir isim olarak kabul edilir.

İçindekiler

Temel varsayımlar ve çerçeve
1. Biyoloji ve sağlık
2. Sinirbilim ve zihin
3. Ekonomik kalkınma ve yoksulluk
4. Barış ve yönetişim
5. İş ve anlam

Ekim 2024

Güçlü yapay zekanın riskleri hakkında çok düşünüyor ve konuşuyorum. CEO'su olduğum şirket, Anthropic, bu riskleri nasıl azaltacağına dair çok fazla araştırma yapıyor. Bu nedenle, insanlar bazen benim yapay zekanın çoğunlukla kötü veya tehlikeli olacağını düşünen bir karamsar veya "kıyametçi" olduğum sonucuna varıyorlar.

Ben hiç öyle düşünmüyorum. Aslında, risklere odaklanmamın ana nedenlerinden biri, bizi temelde olumlu bir gelecek olarak gördüğüm şeyle ayıran tek şeyin onlar olmasıdır. Çoğu insanın yapay zekanın olumlu yönlerinin ne kadar radikal olabileceğini hafife aldığını düşünüyorum, tıpkı çoğu insanın risklerin ne kadar kötü olabileceğini hafife aldığını düşündüğüm gibi.

Bu yazıda, bu olumlu yönlerin nasıl görünebileceğini, her şey yolunda giderse güçlü yapay zekaya sahip bir dünyanın nasıl görünebileceğini ana hatlarıyla çizmeye çalışıyorum. Elbette kimse geleceği kesin veya hassas bir şekilde bilemez ve güçlü yapay zekanın etkileri muhtemelen geçmişteki teknolojik değişikliklerden bile daha tahmin edilemez olacaktır, bu nedenle tüm bunlar kaçınılmaz olarak tahminlerden oluşacaktır. Ancak ben en azından eğitimli ve faydalı tahminler yapmayı hedefliyorum; bu tahminler, çoğu ayrıntı yanlış çıksa bile, olacakların özünü yakalıyor. Birçok ayrıntıyı dahil ediyorum, çünkü bence somut bir vizyon, son derece belirsiz ve soyut bir vizyondan daha fazla tartışmayı ilerletiyor.

Ancak öncelikle, ben ve Anthropic'in neden güçlü yapay zekanın avantajları hakkında çok fazla konuşmadığımızı ve neden genel olarak riskler hakkında çok fazla konuşmaya devam edeceğimizi kısaca açıklamak istedim. Özellikle, bu seçimi şu arzuyla yaptım: Kaldıraç etkisini en üst düzeye çıkarmak. Yapay zeka teknolojisinin temel gelişimi ve birçok (hepsi değil) faydası kaçınılmaz görünüyor (riskler her şeyi raydan çıkarmadığı sürece) ve temelde piyasa güçleri tarafından yönlendiriliyor.

Öte yandan, riskler önceden belirlenmiş değildir ve eylemlerimiz olasılıklarını büyük ölçüde değiştirebilir.

Propaganda algısından kaçınmak. Yapay zeka şirketlerinin yapay zekanın tüm inanılmaz faydalarından bahsetmesi, propagandacı gibi veya olumsuz yönlerden dikkat dağıtmaya çalışıyorlarmış gibi algılanabilir. Ayrıca, prensip olarak, çok fazla zamanınızı "kitabınızı anlatarak" geçirmenin ruhunuz için kötü olduğunu düşünüyorum.

Büyüklenmecilikten kaçınmak. Birçok yapay zeka risk kamu figürünün (yapay zeka şirket liderlerinden bahsetmiyorum bile) yapay zeka sonrası dünyadan bahsetme biçiminden sık sık rahatsız oluyorum; sanki bunu tek başlarına gerçekleştirmek onların göreviymiş gibi, tıpkı bir peygamberin halkını kurtuluşa götürmesi gibi. Bence şirketleri dünyayı tek taraflı olarak şekillendiren varlıklar olarak görmek tehlikeli ve pratik teknolojik hedeflere esasen dini terimlerle bakmak da tehlikeli.

“Bilim kurgu” yükünden kaçının. Çoğu insanın güçlü yapay zekanın avantajlarını hafife aldığını düşünsem de, radikal yapay zeka geleceklerini tartışan küçük bir topluluk bunu genellikle aşırı derecede “bilim kurgu” bir tonda yapıyor (örneğin yüklenmiş zihinler, uzay keşfi veya

genel siberpunk havası[1]). Bence bu, insanların iddiaları daha az ciddiye almasına ve onlara bir tür gerçek dışılık yüklemesine neden oluyor. Açık olmak gerekirse, mesele, açıklanan teknolojilerin mümkün olup olmadığı veya muhtemel olup olmadığı değil. (ana makale bunu ayrıntılı olarak ele alıyor)—mesele daha çok "atmosferin" ima yoluyla bir sürü kültürel yükü ve istenilen geleceğin nasıl olması gerektiği, çeşitli toplumsal sorunların nasıl çözüleceği vb. hakkındaki belirtilmemiş varsayımları içeri sokmasıdır. Sonuç genellikle dar bir alt kültür için bir fantezi gibi okunurken, çoğu insan için itici oluyor.

Yine de yukarıdaki tüm endişelere rağmen, sözü edilen tuzaklardan kaçınmak için elimizden gelenin en iyisini yaparken, güçlü yapay zekâya sahip iyi bir dünyanın nasıl görünebileceğini tartışmanın gerçekten önemli olduğunu düşünüyorum. Aslında bence geleceğe dair gerçekten ilham verici bir vizyona sahip olmak çok önemli; sadece yangın söndürme planı değil. Güçlü yapay zekanın birçok sonucu düşmanca veya tehlikeli olabilir, ancak sonuçta uğruna savaştığımız bir şey olmalı, herkesin daha iyi durumda olduğu, insanları çekişmelerinin üstesinden gelmeye ve önümüzdeki zorluklarla yüzleşmeye teşvik edecek bir şey olmalı. Korku bir tür motivasyon kaynağıdır, ancak yeterli değildir: umuda da ihtiyacımız var.

---

1 Siberpunk, distopik bir geleceğe odaklanan bir fantezi türüdür. Gelişmiş teknoloji, yapay zeka ve gözetleme sistemleriyle şekillenen kentlerde geçen hikâyeler, bireyin özgürlüğü ve mahremiyetini koruma mücadelesini ele alır. Özellikle hükümetler ve dev şirketlerin kontrolü altındaki toplumlarda, bireylerin direnişi temel konular arasındadır. CD PROJEKT RED'in geliştirdiği Cyberpunk 2077 gibi oyunlar, bu dünyayı detaylı bir şekilde canlandırmıştır. Siberpunk, yalnızca eğlence değil, teknolojinin toplumsal etkileri üzerine derinlemesine düşünmeyi teşvik eder.

Güçlü yapay zekanın olumlu uygulamalarının listesi son derece uzundur (ve robotik, üretim, enerji ve daha fazlasını içerir), ancak ben bana göre insan yaşam kalitesini doğrudan iyileştirme potansiyeli en yüksek olan birkaç alana odaklanacağım. En çok heyecan duyduğum beş kategori şunlardır:

1. Biyoloji ve fiziksel sağlık
2. Sinirbilim ve ruh sağlığı
3. Ekonomik kalkınma ve yoksulluk
4. Barış ve yönetişim
5. İş ve anlam

Tahminlerim doğru. Çoğu standarda göre radikal olarak değerlendirilen (bilim kurgu "tekillik" vizyonları hariç), ama bunları içtenlikle ve samimiyetle söylüyorum. Söylediğim her şey çok kolay yanlış olabilir (yukarıdaki duruşumu tekrarlamak gerekirse), ancak en azından görüşlerimi çeşitli alanlardaki ilerlemenin ne kadar hızlanabileceğine ve bunun pratikte ne anlama gelebileceğine dair yarı analitik bir değerlendirmeye dayandırmaya çalıştım.

Hem biyoloji hem de nörobilim alanlarında profesyonel deneyime sahip olduğum için şanslıyım ve ekonomik kalkınma alanında bilgili bir amatörüm, ancak birçok şeyi yanlış söyleyeceğime eminim.

Bu makaleyi yazmanın bana fark ettirdiği bir şey, burada ürettiğimden çok daha iyi ve daha bilgili bir versiyonunu yazmak için bir grup alan uzmanını (biyoloji, ekonomi, uluslararası ilişkiler ve diğer alanlarda) bir araya getirmenin değerli olacağıdır.

Buradaki çabalarımı muhtemelen bu grup için bir başlangıç noktası olarak görmek en iyisidir.

Temel varsayımlar ve çerçeve

Bu makalenin tamamını daha kesin ve sağlam temellere oturtmak için, güçlü yapay zekâdan ne kastettiğimizi (yani 5-10 yıllık sürenin saymaya başladığı eşiği) açıkça belirtmek ve bu tür bir yapay zekânın ortaya çıktıktan sonraki etkilerini düşünmek için bir çerçeve oluşturmak faydalı olacaktır.

Güçlü yapay zekânın (AGI terimini sevmiyorum) neye benzeyeceği ve ne zaman (veya gelip gelmeyeceği) başlı başına büyük bir konu. Bunu kamuoyunda tartıştım ve tamamen ayrı bir makale yazabilirim (muhtemelen bir noktada yazacağım). Açıkçası, birçok insan güçlü yapay zekânın yakında inşa edileceğinden şüphe duyuyor ve bazıları da hiç inşa edilmeyeceğinden şüphe duyuyor.

Bence 2026 gibi erken bir tarihte gelebilir, ancak çok daha uzun sürebilecek yollar da var. Ancak bu denemenin amacı doğrultusunda, bu konuları bir kenara bırakmak, makul bir süre içinde gerçekleşeceğini varsaymak ve ondan sonraki 5-10 yılda neler olacağına odaklanmak istiyorum. Ayrıca, böyle bir sistemin nasıl görüneceğine, yeteneklerinin neler olduğuna ve nasıl etkileşimde bulunduğuna dair bir tanım olduğunu varsaymak istiyorum, her ne kadar bu konuda görüş ayrılığı olsa da.

Güçlü yapay zekâ derken, bugünün LLM'lerine biçim olarak benzer, ancak farklı bir mimariye dayalı olabilecek, birkaç etkileşimli model içerebilecek ve farklı şekilde eğitilebilecek bir yapay zekâ modelini kastediyorum: şu özelliklere sahip: Saf zekâ açısından, en ilgili alanlarda (biyoloji, programlama, matematik, mühendislik, yazarlık vb.) bir Nobel Ödülü sahibinden daha zekidir.

Bu, çözülmemiş matematiksel teoremleri kanıtlayabileceği, son derece iyi romanlar yazabileceği, zor kod tabanlarını sıfırdan yazabileceği vb. anlamına gelir.

Sadece "konuştuğunuz akıllı bir şey" olmanın yanı sıra, sanal olarak çalışan bir insanın sahip olduğu tüm "arayüzlere" sahiptir; metin, ses, video, fare ve klavye kontrolü ve internet erişimi dahil. Bu arayüz tarafından etkinleştirilen her türlü eyleme, iletişime veya uzaktan işleme katılabilir; internette işlem yapmak, insanlara talimat vermek veya onlardan talimat almak, malzeme sipariş etmek, deneyleri yönetmek, videolar izlemek, videolar yapmak vb. Bütün bu görevleri, yine dünyanın en yetenekli insanlarından bile üstün bir beceriyle yapıyor.

Sadece pasif bir şekilde soruları yanıtlamıyor; bunun yerine, tamamlanması saatler, günler veya haftalar süren görevler verilebiliyor ve ardından akıllı bir çalışanın yapacağı gibi, gerektiğinde açıklama isteyerek bu görevleri özerk bir şekilde yapıyor.

Fiziksel bir somutlaşması yok (bilgisayar ekranında yaşaması dışında), ancak mevcut fiziksel araçları, robotları veya laboratuvar ekipmanlarını bir bilgisayar aracılığıyla kontrol edebiliyor; teorik olarak kendi kullanımı için robotlar veya ekipmanlar bile tasarlayabiliyor.

Modelin eğitilmesinde kullanılan kaynaklar, milyonlarca örneğini çalıştırmak için yeniden kullanılabilir (bu, yaklaşık 2027 yılına kadar öngörülen küme boyutlarıyla eşleşir) ve model, bilgiyi emebilir ve eylemleri insan hızının yaklaşık 10 ila 100 katı hızda üretebilir. Ancak, fiziksel dünyanın veya etkileşimde bulunduğu yazılımın tepki süresiyle sınırlı olabilir.

Bu milyon kopyanın her biri, bağımsız olarak ilgisiz görevlerde hareket edebilir veya gerekirse,

insanların iş birliği yapacağı şekilde birlikte çalışabilirler; belki de farklı alt popülasyonlar belirli görevlerde özellikle iyi olacak şekilde ince ayarlanmıştır.

Bunu "veri merkezindeki dâhiler ülkesi" olarak özetleyebiliriz. Açıkçası, böyle bir varlık çok zor problemleri çok hızlı bir şekilde çözebilir, ancak ne kadar hızlı olduğunu anlamak kolay değildir. İki "aşırı" pozisyonun ikisi de bana yanlış görünüyor. İlk olarak, dünyanın saniyeler veya günler ölçeğinde anında dönüşeceğini (“Tekillik”) düşünebilirsiniz, çünkü üstün zekâ kendi üzerine inşa edilerek her türlü bilimsel, mühendislik ve operasyonel görevi neredeyse anında çözecektir. Bununla ilgili sorun şu ki, gerçekte böyle bir durum söz konusu değil. Örneğin donanım yapımı veya biyolojik deneyler yürütme gibi fiziksel ve pratik sınırlamalar. Hatta yeni bir dâhiler ülkesi bile bu sınırlamalara takılacaktır. Zeka çok güçlü olabilir, ama sihirli bir toz değildir.

İkincisi ve bunun tersine, teknolojik ilerlemenin gerçek dünya verileri veya sosyal faktörler tarafından doygunluğa ulaştığına veya hız sınırlamasına uğradığına ve insan zekasından daha iyi zekanın çok az şey katacağına inanabilirsiniz. Bu bana da aynı derecede akıl dışı geliyor—çok zeki bir grup insanın ilerlemeyi önemli ölçüde hızlandırabileceği yüzlerce bilimsel veya hatta sosyal sorun düşünebiliyorum, özellikle de analizle sınırlı değillerse ve gerçek dünyada bir şeyler başarabiliyorlarsa (ki varsayımsal dâhiler ülkemiz bunu yapabilir, insan ekiplerini yönlendirmek veya onlara yardımcı olmak da dahil).

Bence gerçek, bu iki uç resmin karmaşık bir karışımıdır; görev ve alana göre değişen ve ayrıntılarında çok incelikli olan bir şeydir. Bu ayrıntıları verimli bir

şekilde düşünmek için yeni çerçevelere ihtiyacımız olduğuna inanıyorum.

Ekonomistler genellikle "üretim faktörlerinden" bahsederler: emek, toprak ve sermaye gibi şeyler. “Emek/toprak/sermayeye ilişkin marjinal getiriler” ifadesi, belirli bir durumda belirli bir faktörün sınırlayıcı olup olmadığı fikrini yakalar – örneğin, bir hava kuvvetinin hem uçaklara hem de pilotlara ihtiyacı vardır ve uçaklarınız yetersizse daha fazla pilot işe almak pek yardımcı olmaz. Yapay zeka çağında, zekâya ilişkin 7 marjinal getiriden bahsetmemiz ve zekâyı tamamlayıcı olan ve zekâ çok yüksek olduğunda sınırlayıcı faktör haline gelen diğer faktörlerin neler olduğunu anlamaya çalışmamız gerektiğine inanıyorum. Bu şekilde düşünmeye alışkın değiliz – “daha zeki olmak bu görevde ne kadar yardımcı oluyor ve hangi zaman diliminde?” diye sormaya alışkın değiliz – ancak çok güçlü bir yapay zekâya sahip bir dünyayı kavramsallaştırmanın doğru yolu gibi görünüyor.

Zekayı sınırlayan veya tamamlayan faktörlerin bir listesi hakkındaki tahminim şunları içeriyor:

Dış dünyanın hızı. Zeki ajanların, işleri başarmak ve aynı zamanda öğrenmek için dünyada etkileşimli olarak çalışması gerekir. Ancak dünya sadece belirli bir hızda hareket eder. Hücreler ve hayvanlar sabit bir hızda hareket eder, bu nedenle üzerlerinde yapılan deneyler belirli bir süre alır ve bu süre azaltılamaz olabilir. Aynı durum donanım, malzeme bilimi, insanlarla iletişim kurmayı içeren her şey ve hatta mevcut yazılım altyapımız için de geçerlidir. Dahası, bilimde genellikle birçok deney ardışık olarak yapılır ve her biri bir öncekinden öğrenir veya üzerine inşa edilir. Bütün bunlar, büyük bir projenin—örneğin kanser tedavisi geliştirmenin—tamamlanma hızının, zekâ artmaya devam etse bile daha fazla

azaltılamayacak indirgenemez bir minimuma sahip olabileceği anlamına gelir.

Veri ihtiyacı. Bazen ham veri eksiktir ve yokluğunda daha fazla zekâ yardımcı olmaz. Günümüzün parçacık fizikçileri çok zekidir ve çok çeşitli teoriler geliştirmişlerdir, ancak aralarında seçim yapacak veriye sahip değillerdir çünkü parçacık hızlandırıcı verileri çok sınırlıdır. Süper zeki olsalardı, belki de daha büyük bir hızlandırıcı yapımını hızlandırmak dışında, önemli ölçüde daha iyi sonuçlar alacakları açık değildir.

İçsel karmaşıklık. Bazı şeyler doğası gereği tahmin edilemez veya kaotiktir ve en güçlü yapay zekâ bile bunları bugün bir insandan veya bilgisayardan önemli ölçüde daha iyi tahmin edemez veya çözemez. Örneğin, inanılmaz derecede güçlü bir yapay zeka bile, kaotik bir sistemde (üç cisim problemi gibi) yalnızca marjinal olarak daha ileriyi tahmin edebilir. Genel durumda, günümüz insanlarına ve bilgisayarlarına kıyasla.

İnsanlardan kaynaklanan kısıtlamalar. Birçok şey, yasaları çiğnemeden, insanlara zarar vermeden veya toplumu bozmadan yapılamaz. Uyumlu bir yapay zeka bu şeyleri yapmak istemez (ve eğer uyumsuz bir yapay zekamız varsa, risklerden bahsetmeye geri döneriz). Birçok insan toplumsal yapısı verimsizdir veya hatta aktif olarak zararlıdır, ancak klinik denemelerdeki yasal gereklilikler, insanların alışkanlıklarını değiştirme istekliliği veya hükümetlerin davranışı gibi kısıtlamalara saygı duyarken değiştirmek zordur. Teknik anlamda iyi çalışan ancak düzenlemeler veya yersiz korkular nedeniyle etkisi önemli ölçüde azalmış ilerlemelere örnek olarak nükleer enerji, süpersonik ışık ve hatta asansörler verilebilir.

Fiziksel yasalar. Bu, ilk noktanın daha sert bir versiyonudur. Kırılmaz gibi görünen belirli fiziksel yasalar vardır. Işıktan daha hızlı seyahat etmek mümkün değildir. Puding karıştırıldıktan sonra tekrar eski haline dönmez. Çipler, güvenilmez hale gelmeden önce santimetrekare başına yalnızca belirli sayıda transistöre sahip olabilir. Hesaplama, silinen her bit için belirli bir minimum enerji gerektirir ve bu da dünyadaki hesaplama yoğunluğunu sınırlar. Zaman ölçeklerine dayalı bir başka ayrım daha var. Kısa vadede katı kısıtlamalar oluşturan şeyler, uzun vadede zekâya daha açık hale gelebilir. Örneğin, zekâ, eskiden canlı hayvan deneyleri gerektiren şeyleri laboratuvar ortamında öğrenmemizi sağlayan yeni bir deneysel paradigma geliştirmek, yeni veriler toplamak için gereken araçları (örneğin daha büyük parçacık hızlandırıcı) inşa etmek veya (etik sınırlar dahilinde) insan kaynaklı kısıtlamaların üstesinden gelmenin yollarını bulmak (örneğin klinik araştırma sistemini iyileştirmeye yardımcı olmak, klinik araştırmaların daha az bürokrasiye sahip olduğu yeni yetki alanları oluşturmaya yardımcı olmak veya insan klinik araştırmalarını daha az gerekli veya daha ucuz hale getirmek için bilimin kendisini geliştirmek) için kullanılabilir.

Dolayısıyla, zekanın başlangıçta diğer üretim faktörleri tarafından büyük ölçüde kısıtlandığı, ancak zamanla zekanın kendisinin diğer faktörlerin etrafından giderek daha fazla dolaştığı bir tablo hayal etmeliyiz; bu faktörler hiçbir zaman tamamen ortadan kalkmasa bile (ve fizik yasaları gibi bazı şeyler mutlaktır)10. Temel soru, tüm bunların ne kadar hızlı ve hangi sırayla gerçekleştiğidir.

Yukarıdaki çerçeveyi göz önünde bulundurarak, giriş bölümünde bahsedilen beş alan için bu soruyu cevaplamaya çalışacağım.

1. *Biyoloji ve Sağlık*

Biyoloji, muhtemelen bilimsel ilerlemenin insan yaşam kalitesini doğrudan ve net bir şekilde iyileştirme potansiyelinin en yüksek olduğu alandır. İnsan yaşam kalitesini doğrudan ve net bir şekilde iyileştirme potansiyeline sahiptir.

Geçtiğimiz yüzyılda, en eski insan hastalıklarından bazıları (çiçek hastalığı gibi) nihayet yenildi, ancak daha birçok hastalık hala mevcut ve bunların yenilmesi muazzam bir insani başarı olacaktır.

Biyolojik bilim, hastalıkları iyileştirmenin ötesinde, prensip olarak sağlıklı insan yaşam süresini uzatarak, kendi biyolojik süreçlerimiz üzerindeki kontrolü ve özgürlüğü artırarak ve şu anda insanlık durumunun değişmez parçaları olarak düşündüğümüz günlük sorunları ele alarak insan sağlığının temel kalitesini iyileştirebilir.

Önceki bölümdeki "sınırlayıcı faktörler" dilinde, zekayı doğrudan biyolojiye uygulamadaki temel zorluklar veri, fiziksel dünyanın hızı ve içsel karmaşıklıktır (aslında, üçü de birbirleriyle ilişkilidir). İnsan kısıtlamaları, klinik denemeler söz konusu olduğunda daha sonraki bir aşamada da rol oynar. Şimdi bunları tek tek ele alalım.

Hücreler, hayvanlar ve hatta kimyasal süreçler üzerindeki deneyler, fiziksel dünyanın hızıyla sınırlıdır: birçok biyolojik protokol, bakteri veya diğer hücrelerin kültürlenmesini veya kimyasal reaksiyonların gerçekleşmesini beklemeyi içerir ve bu bazen günler hatta haftalar sürebilir, ve bunu hızlandırmanın belirgin bir yolu yoktur. Hayvan deneyleri aylar (veya daha fazla) sürebilir ve insan deneyleri genellikle yıllar (hatta uzun vadeli sonuç çalışmaları için on yıllar) sürer. Bununla biraz bağlantılı olarak, veriler sıklıkla eksiktir—miktar olarak değil, nitelik olarak: İlgilenilen bir biyolojik etkiyi diğer 10.000 karıştırıcı şeyden ayıran, veya belirli bir sürece nedensel olarak müdahale eden veya doğrudan ölçen

(sonuçlarını dolaylı veya gürültülü bir şekilde çıkarmak yerine) açık, net veriler her zaman yetersizdir. Hatta kütle spektrometrisi teknikleri üzerinde çalışırken topladığım proteomik veriler gibi büyük, nicel moleküler veriler bile gürültülüdür ve birçok şeyi kaçırır (bu proteinler hangi hücre tiplerindeydi? Hücrenin hangi bölümünde? Hücre döngüsünün hangi aşamasında?).

Verilerle ilgili bu sorunların kısmen sorumlusu, içsel karmaşıklıktır: İnsan metabolizmasının biyokimyasını gösteren bir diyagram gördüyseniz, bu karmaşık sistemin herhangi bir parçasının etkisini izole etmenin çok zor olduğunu ve sisteme kesin veya tahmin edilebilir bir şekilde müdahale etmenin daha da zor olduğunu bilirsiniz. Son olarak, insanlar üzerinde bir deney yürütmenin gerektirdiği içsel zamanın ötesinde, gerçek klinik denemeler (ben de dahil olmak üzere birçok kişinin görüşüne göre) gereksiz ek zaman ekleyen ve ilerlemeyi geciktiren birçok bürokrasi ve düzenleyici gereklilik içerir.

Tüm bunlar göz önüne alındığında, birçok biyolog uzun zamandır yapay zekanın ve genel olarak "büyük verinin" biyolojideki değerine şüpheyle yaklaşıyor. Tarihsel olarak, son 30 yıldır becerilerini biyolojiye uygulayan matematikçiler, bilgisayar bilimciler ve fizikçiler oldukça başarılı oldular, ancak başlangıçta umulan gerçekten dönüştürücü etkiye sahip olamadılar.

Bazı şüpheler, AlphaFold (ki bu buluş, yaratıcılarına haklı olarak Kimya dalında 11. Nobel Ödülü'nü kazandırdı) veAlphaProteo gibi büyük ve devrim niteliğindeki atılımlarla azaldı, ancak yapay zekanın yalnızca sınırlı bir dizi koşulda yararlı olduğu (ve olmaya devam edeceği) algısı hala mevcut. Yaygın bir ifade şudur: “Yapay zeka verilerinizi daha iyi analiz edebilir, ancak

daha fazla veri üretemez veya verilerin kalitesini iyileştiremez. Çöp girerse, çöp çıkar.”

Ancak bence bu karamsar bakış açısı yapay zekayı yanlış şekilde ele alıyor. Eğer yapay zekânın ilerlemesi hakkındaki temel hipotezimiz doğruysa, o zaman yapay zekâyı düşünmenin doğru yolu, bir veri analiz yöntemi olarak değil, biyologların yaptığı tüm görevleri yerine getiren sanal bir biyolog olarak düşünmektir; bunlar arasında gerçek dünyada deneyler tasarlamak ve yürütmek (laboratuvar robotlarını kontrol ederek veya insanlara hangi deneyleri yapacaklarını söyleyerek - tıpkı bir Baş Araştırmacının yüksek lisans öğrencilerine yaptığı gibi), yeni biyolojik yöntemler veya ölçüm teknikleri icat etmek vb. yer alır. Yapay zekâ, tüm araştırma sürecini hızlandırarak biyolojiyi gerçekten hızlandırabilir. Bunu tekrarlamak istiyorum çünkü yapay zekânın biyolojiyi dönüştürme yeteneğinden bahsettiğimde ortaya çıkan en yaygın yanlış anlama bu: Yapay zekâdan sadece veri analiz aracı olarak bahsetmiyorum. Bu makalenin başındaki güçlü yapay zekâ tanımına uygun olarak, yapay zekâyı biyologların yaptığı hemen hemen her şeyi gerçekleştirmek, yönlendirmek ve geliştirmek için kullanmaktan bahsediyorum.

Hızlanmanın nereden kaynaklanabileceğine dair daha spesifik bir açıklama yapmak gerekirse, biyolojideki ilerlemenin şaşırtıcı derecede büyük bir kısmı, genellikle biyolojik sistemlerde hassas ancak genelleştirilmiş veya programlanabilir müdahaleye olanak sağlayan geniş ölçüm araçları veya teknikleriyle ilgili olan gerçekten çok az sayıda keşiften kaynaklanmaktadır. Genellikle geniş kapsamlı ölçüm araçları veya teknikleriyle ilgili olan biyolojik sistemlerde hassas ancak genelleştirilmiş veya programlanabilir müdahaleye olanak sağlayan keşifler. Yılda belki de bu türden yaklaşık 1 büyük keşif

gerçekleşiyor ve bunların toplu olarak biyolojideki ilerlemenin %50'sinden fazlasını sağladığı söylenebilir. Bu keşifler çok güçlü çünkü tam olarak içsel karmaşıklığı ve veri sınırlamalarını ortadan kaldırarak, biyolojik süreçler üzerindeki anlayışımızı ve kontrolümüzü doğrudan artırıyorlar. Her on yılda bir yapılan birkaç keşif, hem biyolojiye dair temel bilimsel anlayışımızın büyük bir bölümünü mümkün kılmış, hem de en güçlü tıbbi tedavilerin birçoğuna yön vermiştir.

Bazı örnekler şunlardır:

CRISPR: Canlı organizmalarda herhangi bir genin canlı olarak düzenlenmesine olanak tanıyan bir teknik (herhangi bir rastgele gen dizisinin herhangi bir başka rastgele diziyle değiştirilmesi). Orijinal teknik geliştirildiğinden beri, belirli hücre tiplerini hedeflemek, doğruluğu artırmak ve yanlış gen düzenlemelerini azaltmak için sürekli iyileştirmeler yapılmıştır; bunların hepsi insanlarda güvenli kullanım için gereklidir.

Olan biteni hassas bir düzeyde izlemek için çeşitli mikroskopi türleri: gelişmiş ışık mikroskopları (çeşitli

floresan teknikleri, özel optikler vb. ile)[2], elektron mikroskopları[3], atomik kuvvet mikroskopları[4] vb.

Genom dizileme ve sentezi, son birkaç on yılda maliyeti birkaç kat azaldı. Bir nöronun üzerine ışık tutarak ateşlenmesini sağlayan optogenetik teknikler.

Prensip olarak her şeye karşı bir aşı tasarlamamıza ve ardından hızla uyarlamamıza olanak tanıyan mRNA aşıları (mRNA aşıları elbette COVID sırasında ünlendi).

Bağışıklık hücrelerinin vücuttan alınmasına ve prensip olarak her şeye saldırmak üzere "yeniden

---

2 *Gelişmiş Işık Mikroskobu*, biyolojik ve malzeme örneklerini yüksek çözünürlük ve kontrast ile gözlemlemek ve analiz etmek için optik mikroskopideki en son tekniklerin kullanımını ifade eder. Bu teknikler, geleneksel ışık mikroskobunun ulaşamadığı hücresel veya hatta moleküler düzeydeki yapıları ve süreçleri bilim insanlarının görselleştirmesine olanak tanır.

3 *Elektron mikroskopları*, nanometre seviyesindeki objeleri yüksek çözünürlük ve büyütmeye imkan tanır. Hızlandırılmış elektronlar kullanılarak görüntü elde edilir. İki temel türü vardır: İletim Elektron Mikroskobu (TEM) ve Taramalı Elektron Mikroskobu (SEM). TEM, iç yapıları incelemek için; SEM ise yüzey detaylarını görselleştirmek için kullanılır. Bu mikroskoplar bilimsel araştırmada, malzeme bilimi ve biyolojide önemli uygulamalar bulur. Elektron mikroskoplarının icadı, 20. yüzyılın başlarında gelişmiştir.

4 *Atomik kuvvet mikroskobu* (AKM), çok yüksek çözünürlüklü bir taramalı sonda mikroskobu (TSM) türüdür ve yüzeylerin nanoskala özelliklerini görüntülemek için kullanılır. Yüksek çözünürlük, yüzeydeki atomik düzeyde değişimleri göstermekte etkilidir. AKM, yükseklik, sürtünme, manyetizma gibi fiziksel özellikleri ölçmek için kullanılabilir. Bu teknoloji, malzeme bilimi, biyoloji ve nanoçip 研发 alanında önemli uygulamalar sunar.

programlanmasına" olanak tanıyan CAR-T gibi hücre terapileri.

Hastalıkların mikrop teorisi veya bağışıklık sistemi ile kanser arasındaki bağlantının farkına varılması gibi kavramsal içgörüler.

Tüm bu teknolojileri listeleme zahmetine giriyorum çünkü onlar hakkında çok önemli bir iddiada bulunmak istiyorum: Bence keşif oranları, çok daha yetenekli, yaratıcı araştırmacılar olsaydı 10 kat veya daha fazla artırılabilirdi. Ya da başka bir deyişle, bu keşiflerin zekâya getirilerinin yüksek olduğunu ve biyoloji ve tıptaki diğer her şeyin çoğunlukla bunlardan kaynaklandığını düşünüyorum.

Neden böyle düşünüyorum? Çünkü "zekaya getiri"yi belirlemeye çalışırken sormayı alışkanlık haline getirmemiz gereken bazı soruların cevapları var.

Birincisi, bu keşifler genellikle çok az sayıda araştırmacı tarafından, genellikle aynı kişiler tarafından tekrar tekrar yapılıyor; bu da rastgele arama değil, beceri gerektiriyor (ikincisi, uzun deneylerin sınırlayıcı faktör olduğunu düşündürebilir). İkincisi, bunlar genellikle yıllar önce "yapılabilirdi": örneğin, CRISPR, bakterilerdeki bağışıklık sisteminin doğal olarak oluşan bir bileşeniydi ve 80'lerden beri biliniyordu, ancak insanların genel gen düzenlemesi için yeniden kullanılabileceğini fark etmesi 25 yıl daha sürdü. Ayrıca, bilimsel topluluktan umut vadeden yönlere yönelik destek eksikliği nedeniyle de çoğu zaman yıllarca gecikirler (bkz. mRNA aşılarının mucidi hakkındaki bu profil; benzer hikayeler çoktur)[5].

5 mRNA aşılarının mucidi, Macaristan doğumlu bilim insanı Katalin Kariko'dur. Onun çalışmaları, mRNA teknolojisinin güvenli ve etkili şekilde kullanılabileceğini kanıtlamıştır. Bu keşif, COVID-19 aşısı geliştirilirken temel bir rol oynamıştır. 2023 yılında Kariko ve meslektaşı Drew Weissman, bu devrim niteliğindeki buluşları nedeniyle Nobel Tıp Ödülü'ne layık

Üçüncüsü, başarılı projeler genellikle büyük fonlarla desteklenen çabalar yerine, başlangıçta umut vadeden olduğunu düşünmedikleri, sonradan akla gelen veya derme çatma projelerdir. Bu, keşifleri yönlendiren şeyin sadece büyük kaynak yoğunlaşması değil, aynı zamanda zekâ olduğunu da gösteriyor.

Son olarak, bu keşiflerin bazılarının "seri bağımlılığı" olsa da (B keşfini yapmak için araçlara veya bilgiye sahip olmak için önce A keşfini yapmanız gerekir) -ki bu da deneysel gecikmelere yol açabilir- çoğu, belki de çoğu, bağımsızdır; yani aynı anda birçok keşif üzerinde paralel olarak çalışılabilir. Hem bu gerçekler, hem de bir biyolog olarak genel deneyimim, bilim insanları daha zeki olsaydı ve insanlığın sahip olduğu engin biyolojik bilgi arasında bağlantılar kurmada daha iyi olsalardı (CRISPR örneğini tekrar düşünün), yüzlerce keşfin daha yapılabileceğini güçlü bir şekilde düşündürüyor. AlphaFold / AlphaProteo'nun önemli sorunları insanlardan çok daha etkili bir şekilde çözmedeki başarısı, on yıllarca süren dikkatlice tasarlanmış fizik modellemesine rağmen, ileriye dönük yolu göstermesi gereken bir ilke kanıtı (dar bir alanda dar bir araçla da olsa) sağlıyor.

Bu nedenle, güçlü bir yapay zekanın bu keşiflerin hızını en az 10 katına çıkarabileceğini tahmin ediyorum. Bu sayede, önümüzdeki 50-100 yıllık biyolojik ilerlemeyi 5-10 yılda elde edebiliriz.

Neden 100 katı değil? Belki mümkün, ancak burada hem seri bağımlılık hem de deney süreleri önem kazanıyor: 1 yılda 100 yıllık ilerleme elde etmek çok şey gerektiriyor. İlk seferde doğru gitmek için, hayvan deneyleri ve mikroskop tasarımı veya pahalı laboratuvar

---

görülmüştür. Kariko'nun uzun yıllar boyu skeptisizme rağmen sürdürdüğü araştırmalar, gelecekte birçok hastalığa karşı aşıların geliştirilmesine kapı açmıştır.

tesisleri gibi şeyler de dahil. Aslında, 5-10 yılda 1000 yıllık ilerleme kaydedebileceğimiz (belki de absürt gelen) fikre açığım, ancak 1 yılda 100 yıllık ilerleme kaydedebileceğimizden çok şüpheliyim. Başka bir ifadeyle, kaçınılmaz bir sürekli gecikme olduğunu düşünüyorum: deneyler ve donanım tasarımı belirli bir "gecikme"ye sahiptir ve mantıksal olarak çıkarılamayan şeyleri öğrenmek için belirli bir "indirgenemez" sayıda tekrarlanması gerekir. Ancak bunun üzerine büyük ölçekli paralellik mümkün olabilir.

Peki ya klinik denemeler? Her ne kadar bunlarla ilişkili çok fazla bürokrasi ve yavaşlama olsa da, gerçek şu ki, yavaşlıklarının büyük bir kısmı (ama kesinlikle hepsi değil!) nihayetinde neredeyse hiç işe yaramayan veya belirsiz şekilde işe yarayan ilaçları titizlikle değerlendirme ihtiyacından kaynaklanıyor. Bu, ne yazık ki günümüzdeki çoğu tedavi için geçerli: ortalama bir kanser ilacı hayatta kalma süresini birkaç ay artırırken, dikkatlice ölçülmesi gereken önemli yan etkileri de var (Alzheimer ilaçları için de benzer bir durum söz konusu).

Bu, (istatistiksel güç elde etmek için) büyük çalışmalara ve düzenleyici kurumların genellikle bürokrasi ve rekabet eden çıkarların karmaşıklığı nedeniyle yapmakta pek iyi olmadığı zorlu ödünleşmelere yol açar.

Bir şey gerçekten iyi çalıştığında, çok daha hızlı ilerler: Hızlandırılmış bir onay süreci vardır ve etki büyüklükleri daha büyük olduğunda onay kolaylığı çok daha fazladır.

COVID için mRNA aşıları 9 ayda onaylandı - normal hızdan çok daha hızlı. Bununla birlikte, bu koşullar altında bile klinik denemeler hala çok yavaş - mRNA aşıları muhtemelen ~2 ayda onaylanmalıydı. Ancak bu tür gecikmeler (bir ilaç için uçtan uca ~1 yıl), büyük paralelleştirme ve çok fazla olmasa da biraz yineleme

ihtiyacı ("birkaç deneme") ile birleştiğinde, 5-10 yıl içinde radikal bir dönüşümle çok uyumludur.

Daha da iyimser bir bakış açısıyla, yapay zeka destekli biyolojik bilim, insanlarda ne olacağını daha doğru tahmin eden daha iyi hayvan ve hücre deney modelleri (hatta simülasyonlar) geliştirerek klinik denemelerdeki yineleme ihtiyacını azaltabilir. Bu, özellikle yaşlanma sürecine karşı ilaç geliştirmede önemli olacaktır; bu süreç on yıllar boyunca devam eder ve daha hızlı bir yineleme döngüsüne ihtiyacımız vardır.

Son olarak, klinik denemeler ve toplumsal engeller konusunda, biyomedikal yeniliklerin bazı yönlerden başarılı bir şekilde uygulanma konusunda alışılmadık derecede güçlü bir geçmişe sahip olduğunu açıkça belirtmekte fayda var; bu durum diğer bazı teknolojilerin aksine geçerlidir.

Girişte de belirtildiği gibi, birçok teknoloji teknik olarak iyi çalışmasına rağmen toplumsal faktörler tarafından engellenmektedir. Bu, yapay zekanın neler başarabileceğine dair karamsar bir bakış açısı önerebilir. Ancak biyomedikal, ilaç geliştirme sürecinin aşırı derecede zahmetli olmasına rağmen, bir kez geliştirildikten sonra genellikle başarılı bir şekilde uygulanıp kullanılması bakımından benzersizdir.

Yukarıdakileri özetlemek gerekirse, temel tahminim yapay zeka destekli biyoloji ve tıbbın, insan biyologlarının önümüzdeki 50-100 yılda elde edeceği ilerlemeyi 5-10 yıla sıkıştırmamıza olanak sağlayacağıdır. Buna “sıkıştırılmış 21. yüzyıl” diyeceğim: güçlü bir yapay zeka geliştirildikten sonra, birkaç yıl içinde tüm 21. yüzyılda yapacağımız biyoloji ve tıp alanındaki tüm ilerlemeyi yapacağımız fikri.

Güçlü bir yapay zekanın birkaç yıl içinde neler yapabileceğini tahmin etmek doğası gereği zor ve

spekülatif olsa da, “insanlar önümüzdeki 100 yılda yardımsız neler yapabilir?” sorusunu sormanın bir somutluğu vardır. Basitçe 20. yüzyılda başardıklarımıza bakmak, 21. yüzyılın ilk 20 yılından yola çıkarak tahminlerde bulunmak veya “10 CRISPR ve 50 CAR-T”nin bize ne kazandıracağını sormak, güçlü yapay zekâdan bekleyebileceğimiz genel ilerleme düzeyini tahmin etmek için pratik ve sağlam temellere dayalı yollar sunar.

Aşağıda, neler bekleyebileceğimize dair bir liste yapmaya çalışıyorum. Bu, herhangi bir titiz metodolojiye dayanmamaktadır ve ayrıntılarda neredeyse kesinlikle yanlış çıkacaktır, ancak beklememiz gereken genel radikallik düzeyini aktarmaya çalışmaktadır: Neredeyse tüm doğal bulaşıcı hastalığın güvenilir bir şekilde önlenmesi ve tedavisi. 20. yüzyılda bulaşıcı hastalıklara karşı elde edilen muazzam ilerlemeler göz önüne alındığında, sıkıştırılmış bir 21. yüzyılda “işi bitirebileceğimizi” hayal etmek radikal değildir. mRNA aşıları ve benzer teknolojiler zaten “her şey için aşılar”a doğru yolu göstermektedir. Bulaşıcı hastalıkların dünyadan tamamen ortadan kaldırılıp kaldırılmayacağı (sadece bazı yerlerde değil), Bölüm 3'te ele alınan yoksulluk ve eşitsizlik sorularına bağlıdır.

Çoğu kanserin ortadan kaldırılması. Kanserden ölüm oranları son birkaç on yıldır yılda yaklaşık %2 oranında düşüyor; bu nedenle, insan biliminin mevcut hızıyla 21. yüzyılda çoğu kanseri ortadan kaldırma yolunda ilerliyoruz. Bazı alt tipler büyük ölçüde iyileştirildi (örneğin, CAR-T terapisi ile bazı lösemi türleri) ve ben belki de kanseri başlangıç aşamasında hedef alan ve büyümesini engelleyen çok seçici ilaçlar için daha da heyecanlıyım.

Yapay zeka ayrıca tedavi rejimlerini kanserin bireyselleştirilmiş genomuna çok ince bir şekilde uyarlamayı mümkün kılacak; bunlar bugün mümkün, ancak zaman ve insan uzmanlığı açısından son derece pahalı, ki yapay zeka bunu ölçeklendirmemize olanak sağlamalı. Hem ölüm oranlarında hem de vaka oranlarında %95 veya daha fazla azalma mümkün görünüyor.

Bununla birlikte, kanser son derece çeşitli ve uyarlanabilir bir hastalıktır ve muhtemelen bu hastalıkların en zor yok edilenidir. Nadir ve zorlu kanser türlerinin varlığını sürdürmesi şaşırtıcı olmazdı.

Genetik hastalıklar için çok etkili önleme ve etkili tedaviler mevcut.

Büyük ölçüde geliştirilmiş embriyo taraması, genetik hastalıkların çoğunu önlemeyi mümkün kılacak ve CRISPR'ın daha güvenli ve daha güvenilir bir türevi, mevcut insanlardaki genetik hastalıkların çoğunu tedavi edebilir. Ancak, hücrelerin büyük bir bölümünü etkileyen tüm vücudu etkileyen rahatsızlıklar son kalanlar olabilir.

Alzheimer'ın önlenmesi. Alzheimer'a neyin neden olduğunu anlamakta çok zorlandık (bir şekilde beta-amiloid proteini ile ilişkili, ancak gerçek ayrıntılar çok karmaşık görünüyor). Biyolojik etkileri izole eden daha iyi ölçüm araçlarıyla çözülebilecek bir sorun gibi görünüyor; bu nedenle yapay zekanın bunu çözme yeteneği konusunda iyimserim. Sonunda nispeten basit müdahalelerle önlenebilme olasılığı oldukça yüksek, ancak olanları gerçekten anladığımızda.

Bununla birlikte, zaten var olan Alzheimer'dan kaynaklanan hasarı tersine çevirmek çok zor olabilir.

Diğer birçok rahatsızlığın tedavisinde iyileşme. Bu, diyabet, obezite, kalp hastalığı, otoimmün hastalıklar ve daha fazlasını içeren diğer rahatsızlıklar için genel bir kategoridir. Bunların çoğu kanser ve Alzheimer'dan daha

"kolay" çözülebilir gibi görünüyor ve birçok durumda zaten büyük bir düşüşte. Örneğin, kalp hastalığından ölümler zaten %50'den fazla azaldı ve GLP-1 agonistleri gibi basit müdahaleler obezite ve diyabete karşı büyük ilerleme kaydetti.

Biyolojik özgürlük. Son 70 yılda doğum kontrolü, doğurganlık, kilo yönetimi ve daha birçok alanda ilerlemeler kaydedildi. Ancak yapay zekanın hızlandırdığı biyolojinin, mümkün olanı büyük ölçüde genişleteceğinden şüpheleniyorum: kilo, fiziksel görünüm, üreme ve diğer biyolojik süreçler tamamen insanların kontrolü altında olacak. Bunlara biyolojik özgürlük başlığı altında değineceğiz: herkesin ne olmak istediğini seçme ve hayatını en çok hoşuna giden şekilde yaşama gücüne sahip olması gerektiği fikri.

Elbette küresel erişim eşitliği hakkında önemli sorular olacaktır; bunlar için Bölüm 3'e bakın. İnsan ömrünün ikiye katlanması. Bu radikal görünebilir, ancak yaşam beklentisi 20. yüzyılda neredeyse 2 katına çıktı (~40 yıldan ~75 yıla), bu nedenle "sıkıştırılmış 21. yüzyılın" bunu tekrar ikiye katlayarak 150 yıla çıkarması "trend"e uygun. Açıkçası, gerçek yaşlanma sürecini yavaşlatmak için gereken müdahaleler, geçen yüzyılda (çoğunlukla çocukluk çağı) hastalıktan kaynaklanan erken ölümleri önlemek için gerekenlerden farklı olacaktır, ancak değişimin büyüklüğü emsalsiz değildir. Somut olarak, halihazırda sınırlı yan etkilerle sıçanlarda maksimum yaşam süresini %25-50 oranında artıran ilaçlar mevcuttur. Ve bazı hayvanlar (örneğin bazı kaplumbağa türleri) zaten 200 yıl yaşıyor, bu nedenle insanların teorik bir üst sınıra ulaşmadığı açıkça ortada.

Tahminime göre, en önemli şey, insan yaşlanmasının güvenilir, Goodhart olmayan

biyobelirteçleri olabilir[6], çünkü bu deneyler ve klinik çalışmalar üzerinde hızlı yinelemelere olanak sağlayacaktır. İnsan ömrü 150 yıla ulaştığında, "kaçış hızına" ulaşabilir ve bugün hayatta olanların çoğunun istedikleri kadar yaşayabilmeleri için yeterli zaman kazanabiliriz, ancak bunun biyolojik olarak mümkün olduğuna dair kesinlikle bir garanti yok[7].

Bu listeye bakmak ve bunların hepsinin 7-12 yıl içinde (ki bu agresif bir yapay zeka zaman çizelgesiyle uyumlu olacaktır) başarılması durumunda dünyanın ne kadar farklı olacağını düşünmekte fayda var.

6 Goodhart olmayan biyobelirteçler, tıpta tanı ve tedavi süreçlerinde objektif ölçülebilir biyolojik göstergeler olarak tanımlanır. Bu biyobelirteçler, hastalığın varlığını, tedavi yanıtını ve ilerlemesini değerlendirmek için kullanılır. Örneğin, kanser biyobelirteçleri, normal dokunun malign olma ihtimalini gösterirken, akut kalp yetersizliğinde kullanılan natriüretik peptitler, hastalığın tanısını ve tedaviyi değerlendirmek için önemli bir rol oynar. Bu biyobelirteçler, hastalıkların teşhisi, önlenmesi ve tedavisi ile ilgili kararlar almak için kullanılır ve kişiselleştirilmiş tıp gibi yeni yaklaşımlarla birlikte tıbbın uygulanma şeklini değiştirebilir.

7 Buradaki “kaçış hızı” kavramı, astronomideki teknik anlamıyla doğrudan aynı şey değil ama bilinçli bir metafor. Astronomideki kaçış hızı Bir cismin, örneğin bir roketin, bir gezegenin veya yıldızın yerçekiminden kurtulabilmesi için ulaşması gereken minimum hızdır. Dünya için bu hız yaklaşık 11,2 km/s’dir. Uzun ömür ve biyolojideki “kaçış hızı” metaforu Yaşlanma araştırmalarında “longevity escape velocity” (uzun ömür kaçış hızı) diye bir kavram var. Anlamı: Tıbbi ve biyoteknolojik ilerlemeler, insanların yaşam süresini her yıl en az bir yıl uzatacak kadar hızlı gelişirse, bireyler sürekli olarak “ölümden kaçabilir.” Yani, bilim her yıl yaşlanmayı yavaşlatma veya tersine çevirme konusunda yeterince ilerleme kaydederse, insanlar teorik olarak sınırsız süre yaşayabilir. Bağıntı Astronomideki kavramdan esinlenilmiş: Yerçekiminden kurtulmak için belli bir hız eşiği gerekir; aynı şekilde yaşlanmadan “kurtulmak” için de bilimsel ilerlemenin belli bir hız eşiğine ulaşması gerekir. Dolayısıyla, biyolojideki kullanım doğrudan fiziksel bir hesap değil, metaforik bir aktarım.

Binlerce yıldır insanlığı rahatsız eden belaların çoğunun bir anda ortadan kaldırılması, hayal edilemez bir insani zafer olacağı aşikardır.

Arkadaşlarımın ve meslektaşlarımın çoğu çocuk yetiştiriyor ve bu çocuklar büyüdüğünde, hastalıktan bahsetmenin onlara iskorbüt, çiçek hastalığı veya veba gibi gelmesini umuyorum. Bu nesil ayrıca artan biyolojik özgürlükten de faydalanacak. Özgürlük ve kendini ifade etme özgürlüğü ve şanslılarsa istedikleri kadar uzun yaşama imkanı da bulabilirler. Bu değişikliklerin, güçlü yapay zekâ bekleyen küçük bir topluluk dışında herkes için ne kadar şaşırtıcı olacağını abartmak zor. Örneğin, ABD'de binlerce ekonomist ve politika uzmanı şu anda Sosyal Güvenlik ve Medicare'i nasıl ayakta tutacaklarını ve daha genel olarak sağlık hizmetlerinin maliyetini nasıl düşüreceklerini tartışıyor (ki bu çoğunlukla 70 yaş üstü kişiler ve özellikle kanser gibi ölümcül hastalıkları olanlar tarafından tüketiliyor). Tüm bunlar gerçekleşirse, bu programların durumu büyük ölçüde iyileşecektir, çünkü çalışma çağındaki nüfusun emekli nüfusa oranı önemli ölçüde değişecektir. Şüphesiz ki bu zorlukların yerini, yeni teknolojilere yaygın erişimi nasıl sağlayacağımız gibi başka zorluklar alacaktır; ancak biyolojinin yapay zekâ tarafından başarıyla hızlandırılan tek alan olsa bile dünyanın ne kadar değişeceği üzerinde düşünmeye değer.

Sinirbilim ve zihin

Önceki bölümde fiziksel hastalıklara ve genel olarak biyolojiye odaklandım ve sinirbilim veya ruh sağlığına değinmedim. Ancak sinirbilim, biyolojinin bir alt disiplinidir ve ruh sağlığı da fiziksel sağlık kadar önemlidir. Aslında, ruh sağlığı insan refahını fiziksel sağlıktan daha doğrudan etkiler. Yüz milyonlarca insan, bağımlılık, depresyon, şizofreni, düşük işlevli otizm, travma sonrası stres bozukluğu, psikopati veya zihinsel

engellilik gibi sorunlar nedeniyle çok düşük yaşam kalitesine sahip.

Milyarlarca insan daha, genellikle bu ciddi klinik bozuklukların çok daha hafif versiyonları olarak yorumlanabilen günlük sorunlarla mücadele ediyor. Ve genel biyolojide olduğu gibi, sorunları ele almanın ötesine geçerek insan deneyiminin temel kalitesini iyileştirmek mümkün olabilir. Biyoloji için ortaya koyduğum temel çerçeve, nörobilim için de aynı şekilde geçerlidir.

Bu alan, genellikle ölçüm veya hassas müdahale araçlarıyla ilgili az sayıda keşif sayesinde ilerlemektedir – yukarıdaki listede optogenetik bir nörobilim keşfiydi ve daha yakın zamanda CLARITY ve genişleme mikroskopisi de aynı doğrultudaki gelişmelerdir; ayrıca birçok genel hücre biyolojisi yöntemi doğrudan nörobilime aktarılmaktadır.

Bence bu gelişmelerin hızı yapay zeka tarafından da benzer şekilde hızlandırılacak ve bu nedenle "5-10 yılda 100 yıllık ilerleme" çerçevesi, biyolojide olduğu gibi nörobilim için de aynı şekilde ve aynı nedenlerle geçerlidir. Biyolojide olduğu gibi, 20. yüzyıl sinirbilimindeki ilerleme de muazzamdı – örneğin, nöronların nasıl veya neden ateşlendiğini 1950'lere kadar anlamıyorduk. Bu nedenle, yapay zekanın hızlandırılmış sinirbiliminin birkaç yıl içinde hızlı ilerleme sağlamasını beklemek makul görünüyor.

Bu temel tabloya eklememiz gereken bir şey daha var: son birkaç yılda yapay zeka hakkında öğrendiğimiz (veya öğrenmekte olduğumuz) bazı şeyler, sadece insanlar tarafından yapılmaya devam etse bile, sinirbilimin ilerlemesine yardımcı olma olasılığı yüksek.

Yorumlanabilirlik açık bir örnektir: Biyolojik nöronlar yüzeysel olarak yapay nöronlardan tamamen farklı bir şekilde çalışsa da (sinyaller ve genellikle sinyal

hızları aracılığıyla iletişim kurarlar, bu nedenle yapay nöronlarda bulunmayan bir zaman unsuru vardır ve hücre fizyolojisi ve nörotransmiterlerle ilgili bir dizi ayrıntı, çalışma şekillerini önemli ölçüde değiştirir), "dağıtılmış, eğitilmiş, birleştirilmiş doğrusal/doğrusal olmayan işlemleri gerçekleştiren basit birim ağları nasıl birlikte çalışarak önemli hesaplamalar yapar" temel sorusu aynıdır ve bireysel nöron iletişiminin ayrıntılarının, hesaplama ve devrelerle ilgili ilginç sorunun çoğunda soyutlanacağından şüpheleniyorum. Bunun sadece bir örneği olarak, yapay zeka sistemlerinde yorumlanabilirlik araştırmacıları tarafından keşfedilen bir hesaplama mekanizması yakın zamanda farelerin beyinlerinde yeniden keşfedildi.

Yapay sinir ağları üzerinde deney yapmak, gerçek ağlar üzerinde deney yapmaktan çok daha kolaydır (ikincisi genellikle hayvan beyinlerine müdahale etmeyi gerektirir), bu nedenle yorumlanabilirlik, sinirbilim anlayışımızı geliştirmek için bir araç haline gelebilir. Ayrıca, güçlü yapay zekâlar muhtemelen bu aracı insanlardan daha iyi geliştirip uygulayabileceklerdir.

Ancak yorumlanabilirliğin ötesinde, yapay zekâdan öğrendiklerimiz, akıllı sistemlerin nasıl eğitildiği konusunda (henüz emin olmasam da) sinirbilimde bir devrime yol açmalıdır.

Sinirbilim alanında çalışırken, birçok insan şimdi yanlış sorular olarak değerlendireceğim öğrenme hakkındaki sorulara odaklanıyordu, çünkü ölçeklendirme hipotezi / acı ders kavramı henüz mevcut değildi[8]. Basit

8 *Ölçeklenme hipotezi*, özellikle makine öğrenimi ve kritik fenomenlerde, sistemlerin boyut veya veri miktarı arttıkça yeni özelliklerin ortaya çıkacağını öne sürer. Prof. Kardar'ın tanımladığı gibi, bu hipotez homojenite varsayımı, korelasyon uzunluğunun diverjansı ve kendine benzerlikle ilişkilidir. Yeni veri ve hesaplama gücüyle birlikte sinir ağları, daha genelleyici ve Bayesian davranışı göstererek zor problemlerde beklenmedik

bir amaç fonksiyonunun artı çok miktarda verinin inanılmaz derecede karmaşık davranışları yönlendirebileceği fikri, amaç fonksiyonlarını ve mimari önyargıları anlamayı daha ilginç, ortaya çıkan hesaplamaların ayrıntılarını anlamayı ise daha az ilginç hale getiriyor.

Son yıllarda bu alanı yakından takip etmedim, ancak hesaplamalı sinirbilimcilerin bu dersi henüz tam olarak özümsemediklerine dair belirsiz bir hissim var.

Ölçekleme hipotezine karşı tutumum her zaman "aha - bu, zekanın nasıl çalıştığının ve nasıl bu kadar kolay evrimleştiğinin yüksek düzeyde bir açıklaması" olmuştur, ancak bunun ortalama bir sinirbilimcinin görüşü olduğunu düşünmüyorum, kısmen de ölçekleme hipotezinin "zekanın sırrı" olarak yapay zeka içinde bile tam olarak kabul görmemesinden kaynaklanıyor.

Bence sinirbilimciler bu temel anlayışı insan beyninin özellikleriyle (biyofiziksel sınırlamalar, evrimsel tarih, topoloji, motor ve duyusal girdi/çıktıların ayrıntıları) birleştirmeye çalışmalı ve sinirbilimin temel bulmacalarından bazılarını çözmeye çalışmalıdır. Bazıları muhtemelen bunu yapıyor, ancak henüz yeterli olmadığını ve yapay zeka sinirbilimcilerinin bu açıdan daha etkili bir şekilde yararlanarak ilerlemeyi hızlandırabileceğini düşünüyorum.

---

yetenekler sergileyebilir. Bu süreç, ölçek büyüdükçe insan düzeyinde yeteneklerin oluştuğunu düşündürür. *Bayesian davranış*, bilişsel süreçlerde incelenen bilgileri yeni verilerle birleştirerek inançları güncelleme yaklaşımıdır. Sinirbilimde, sinir hücrelerinin davranış kalıplarını tahmin etmek için Bayes teoremi kullanılır; bu, gözlemlenen hareketlerle beynin işleyişini ilişkilendirmede etkilidir. Bayes istatistikleri, olasılığı bir inanç düzeyi olarak görerek, parametreler üzerine doğrudan olasılık dağılımları atayabilir. Bu, nöroprotez ve beyin-bilgisayar arayüzleri gibi alanlarda davranış tahmini ve gerçek zamanlı uyum sağlama açısından önemli avantajlar sunar.

Yapay zekanın sinirbilimsel ilerlemeyi dört farklı yolda hızlandıracağını ve bunların hepsinin umarım zihinsel hastalıkları tedavi etmek ve işlevselliği iyileştirmek için birlikte çalışabileceğini düşünüyorum: Geleneksel moleküler biyoloji, kimya ve genetik. Bu, esas olarak 1. bölümdeki genel biyolojiyle aynı hikaye ve yapay zeka muhtemelen aynı mekanizmalar aracılığıyla bunu hızlandırabilir. Beyin fonksiyonunu değiştirmek, uyanıklığı veya algıyı etkilemek, ruh halini değiştirmek vb. için nörotransmitterleri modüle eden birçok ilaç var ve yapay zeka bize çok daha fazlasını icat etmemize yardımcı olabilir. Yapay zeka muhtemelen zihinsel hastalıkların genetik temeli üzerine yapılan araştırmaları da hızlandırabilir.

İyi-Tanecikli sinirsel ölçüm ve müdahale. Bu, birçok bireysel nöronun veya nöronal devrenin ne yaptığını ölçme ve davranışlarını değiştirmek için müdahale etme yeteneğidir. Optogenetik ve sinirsel problar, canlı organizmalarda hem ölçüm hem de müdahale yapabilen teknolojilerdir ve bir dizi çok gelişmiş yöntem (örneğin, çok sayıda bireysel nöronun ateşleme modellerini okumak için moleküler bantlar) önerilmiştir ve prensipte mümkün görünmektedir. Ayrıca gelişmiş hesaplamalı sinirbilim de mevcuttur. Yukarıda belirtildiği gibi, hem spesifik içgörüler hem de modern yapay zekanın bütünlüğü, muhtemelen sistem sinirbilimindeki sorulara verimli bir şekilde uygulanabilir; örneğin, psikoz veya duygu bozuklukları gibi karmaşık hastalıkların gerçek nedenlerini ve dinamiklerini ortaya çıkarmak gibi.

Davranışsal müdahaleler. Sinirbilimin biyolojik yönüne odaklanıldığı için bundan pek bahsetmedim, ancak psikiyatri ve psikoloji elbette 20. yüzyıl boyunca geniş bir davranışsal müdahale repertuarı geliştirdi; yapay zekanın da hem yeni yöntemlerin geliştirilmesini hem de hastaların

mevcut yöntemlere uymasına yardımcı olarak bunları hızlandırabileceği mantıklı geliyor. Daha genel olarak, her zaman kendinizin en iyi versiyonu olmanıza yardımcı olan, etkileşimlerinizi inceleyen ve daha etkili olmayı öğrenmenize yardımcı olan bir "yapay zeka koçu" fikri çok umut verici görünüyor.

Tahminimce, bu dört ilerleme yolu birlikte çalışırsa, tıpkı fiziksel hastalıklarda olduğu gibi, yapay zekâ devreye girmese bile önümüzdeki 100 yıl içinde çoğu ruhsal hastalığın tedavisine veya önlenmesine yol açacaktır ve bu nedenle yapay zekânın hızlandırdığı 5-10 yıl içinde makul bir şekilde tamamlanabilir. Somut olarak, olacaklar hakkındaki tahminim şöyle bir şey: Çoğu akıl hastalığı muhtemelen tedavi edilebilir. Psikiyatrik hastalıklar konusunda uzman değilim (nörobilimdeki zamanımı küçük nöron gruplarını incelemek için problar inşa ederek geçirdim), ancak tahminimce travma sonrası stres bozukluğu, depresyon, şizofreni, bağımlılık vb. hastalıklar yukarıdaki dört yönün bir kombinasyonu yoluyla çok etkili bir şekilde tedavi edilebilir.

Cevap muhtemelen "biyokimyasal olarak bir şeyler ters gitti" (çok karmaşık olsa da) ve "sinir ağında, üst düzeyde bir şeyler ters gitti"nin bir kombinasyonu olacaktır. Yani, bu bir sistem nörobilimi sorusudur; ancak bu, yukarıda tartışılan davranışsal müdahalelerin etkisini inkar etmez. Ölçüm ve müdahale araçları, özellikle canlı insanlarda, hızlı yinelemeye ve ilerlemeye yol açacak gibi görünüyor.

Çok "yapısal" olan durumlar daha zor olabilir, ancak imkansız değildir. Psikopatinin belirgin nöroanatomik farklılıklarla ilişkili olduğuna dair bazı kanıtlar var; bazı beyin bölgeleri psikopatlarda daha küçük veya daha az gelişmiştir. Psikopatların genç yaştan itibaren empati yoksunu olduğuna da inanılıyor; beyinlerinde farklı

olan ne olursa olsun, muhtemelen her zaman böyleydi. Aynı şey bazı zihinsel engellilikler ve belki de diğer durumlar için de geçerli olabilir. Beyni yeniden yapılandırmak zor görünüyor, ancak zekâ açısından yüksek getiri sağlayan bir görev gibi de görünüyor. Belki yetişkin beynini daha erken veya daha esnek bir duruma getirmenin bir yolu vardır, böylece yeniden şekillendirilebilir. Bunun ne kadar mümkün olduğundan çok emin değilim, ancak içgüdüm yapay zekanın burada neler icat edebileceği konusunda iyimser olmamı söylüyor.

Zihinsel hastalıkların etkili genetik önlenmesi mümkün görünüyor. Çoğu zihinsel hastalık kısmen kalıtsaldır ve genom çapında ilişki çalışmaları, genellikle çok sayıda olan ilgili faktörleri belirleme konusunda ivme kazanmaya başlıyor. Muhtemelen bu hastalıkların çoğunu embriyo taraması yoluyla önlemek mümkün olacak, fiziksel hastalıklarla ilgili hikayeye benzer şekilde. Bir fark şudur ki, psikiyatrik hastalıkların poligenik olma olasılığı daha yüksektir (birçok gen katkıda bulunur), bu nedenle karmaşıklık nedeniyle, hastalıkla ilişkili olumlu özelliklere karşı bilmeden seçim yapma riski artar. Ancak garip bir şekilde, son yıllarda GWAS çalışmaları bu korelasyonların abartılmış olabileceğini öne sürüyor gibi görünüyor[9]. Her durumda, yapay zeka destekli sinirbilim bu şeyleri anlamamıza yardımcı olabilir. Elbette, karmaşık özellikler için embriyo taraması bir dizi toplumsal sorunu

9 Genom genişliği ilişkilendirme çalışması (GWAS), genomdaki genetik varyasyonların belirli hastalıklar veya fenotiplerle bağlantılı olup olmadığını belirlemek için kullanılan güçlü bir yöntemdir. NHGRI-EBI GWAS Kataloğu, tüm insan GWAS çalışmalarının kapsamlı ve dikkatlice derlenen bir koleksiyonudur. Bu çalışma, kalıtsal hastalıkların temel mekanizmalarını anlamada kritik rol oynar ve yeni hedeflerin tespitinde yardımcı olur. Varyasyonlar, özellikle tek nükleotid polimorfizmaları (SNP'ler) gibi küçük değişiklikler, hastalık riski ile istatistiksel olarak ilişkilendirilir.

gündeme getiriyor ve tartışmalı olacak, ancak sanırım çoğu insan ciddi veya sakatlayıcı zihinsel hastalıklar için taramayı destekleyecektir.

Klinik hastalık olarak düşünmediğimiz günlük sorunlar da çözülecektir. Çoğumuzun günlük psikolojik sorunları vardır ve bunlar genellikle klinik hastalık düzeyine yükseldiği düşünülmez. Bazı insanlar çabuk sinirlenir, bazılarının odaklanma sorunu vardır veya sık sık uykuludur, bazıları korkulu veya endişelidir veya değişime kötü tepki verir. Günümüzde, örneğin uyanıklık veya odaklanma konusunda yardımcı olacak ilaçlar zaten mevcuttur (kafein, modafinil, (ritalin) ancak diğer birçok önceki alanda olduğu gibi, çok daha fazlasının mümkün olması muhtemel. Muhtemelen daha birçok ilaç mevcut ve henüz keşfedilmemiş durumda ve ayrıca hedefli ışık uyarımı (yukarıdaki optogenetiğe bakın) veya manyetik alanlar gibi tamamen yeni müdahale yöntemleri de olabilir.

20. yüzyılda bilişsel işlevi ve duygusal durumu ayarlayan ne kadar çok ilaç geliştirdiğimizi göz önünde bulundurarak, herkesin beyninin biraz daha iyi davranmasını ve daha tatmin edici bir günlük deneyim yaşamasını sağlayabileceği "sıkıştırılmış 21. yüzyıl" konusunda çok iyimserim.

İnsan temel deneyimi çok daha iyi olabilir. Bir adım daha ileri giderek, birçok insan olağanüstü anlar yaşadı: vahiy, yaratıcı ilham, şefkat, tatmin, aşk, güzellik veya meditatif huzur. Bu deneyimlerin niteliği ve sıklığı kişiden kişiye ve aynı kişide farklı zamanlarda büyük ölçüde değişir ve ayrıca bazen çeşitli ilaçlar tarafından tetiklenebilir (çoğu zaman yan etkilerle birlikte). Bütün bunlar, “deneyimlenebilecek şeylerin alanı”nın çok geniş olduğunu ve insanların yaşamlarının daha büyük bir bölümünün bu olağanüstü anlardan oluşabileceğini

düşündürmektedir. Muhtemelen genel olarak çeşitli bilişsel işlevleri iyileştirmek de mümkündür. Bu belki de "biyolojik özgürlük" veya "uzun ömür"ün nörobilim versiyonudur.

Yapay zekanın bilim kurgu tasvirlerinde sıklıkla ortaya çıkan, ancak burada kasıtlı olarak ele almadığım bir konu da "zihin yükleme", yani insan beyninin kalıbını ve dinamiklerini yakalama ve bunları yazılımda somutlaştırma fikridir. Bu konu başlı başına bir makalenin konusu olabilir, ancak şunu söylemek yeterli olacaktır ki, yüklemenin prensipte neredeyse kesinlikle mümkün olduğunu düşünsem de, pratikte, güçlü yapay zekâ ile bile, önemli teknolojik ve toplumsal zorluklarla karşı karşıyadır ve bu da onu muhtemelen tartıştığımız 5-10 yıllık zaman diliminin dışında bırakacaktır.

Özetle, yapay zekâ ile hızlandırılmış sinirbilim, çoğu akıl hastalığının tedavisini büyük ölçüde iyileştirecek, hatta iyileştirecek ve ayrıca "bilişsel ve zihinsel özgürlüğü" ve insan bilişsel ve duygusal yeteneklerini büyük ölçüde genişletecektir. Önceki bölümde açıklanan fiziksel sağlık alanındaki iyileştirmeler kadar radikal olacaktır. Belki de dünya dışarıdan görünürde farklı olmayacak, ancak insanların deneyimlediği dünya çok daha iyi ve daha insancıl bir yer olacak, aynı zamanda kendini gerçekleştirme için daha büyük fırsatlar sunan bir yer olacak. Ayrıca, gelişmiş ruh sağlığının, siyasi veya ekonomik görünenler de dahil olmak üzere birçok diğer toplumsal sorunu iyileştireceğinden şüpheleniyorum.

Ekonomik kalkınma ve yoksulluk

Önceki iki bölüm, hastalığı iyileştiren ve insan yaşam kalitesini artıran yeni teknolojilerin geliştirilmesiyle ilgilidir. Ancak insani bir bakış açısından

açık bir soru şudur: "Herkes bu teknolojilere erişebilecek mi?"

Bir hastalığa çare geliştirmek bir şeydir, hastalığı dünyadan tamamen ortadan kaldırmak ise başka bir şeydir. Daha genel olarak, mevcut birçok sağlık müdahalesi henüz dünyanın her yerinde uygulanmamıştır ve bu durum (sağlık dışı) teknolojik gelişmeler için de geçerlidir. Bunu başka bir şekilde ifade etmek gerekirse, dünyanın birçok yerinde yaşam standartları hala son derece düşüktür: Sahra Altı Afrika'da kişi başına GSYİH yaklaşık 2.000 dolar iken, Amerika Birleşik Devletleri'nde yaklaşık 75.000 dolardır. Yapay zeka, gelişmiş dünyada ekonomik büyümeyi ve yaşam kalitesini daha da artırırken, gelişmekte olan dünyaya pek bir şey yapmazsa, bunu korkunç bir ahlaki başarısızlık ve önceki iki bölümdeki gerçek insani zaferlere bir leke olarak görmeliyiz. İdeal olarak, güçlü yapay zeka, gelişmekte olan dünyanın gelişmiş dünyaya yetişmesine yardımcı olmalı, hatta ikincisini devrimleştirmelidir.

Yapay zekanın eşitsizliği ve ekonomik büyümeyi ele alabileceğinden, temel teknolojileri icat edebileceğinden daha az eminim, çünkü teknoloji, zekaya (karmaşıklıkların ve veri eksikliğinin etrafından dolaşma yeteneği de dahil olmak üzere) çok açık bir şekilde yüksek getiri sağlarken, ekonomi insanlardan kaynaklanan birçok kısıtlamanın yanı sıra büyük bir içsel karmaşıklık dozu da içeriyor. Bir yapay zekanın ünlü "sosyalist hesaplama problemini" çözebileceğinden biraz şüpheliyim[10] ve ve

10 Sosyalist hesaplama sorunu, sosyalist ekonomilerin kaynak dağılımını etkin bir şekilde yönetebilmesi konusundaki teorik zorlukları ifade eder. Ludwig von Mises ve Friedrich Hayek'e göre, serbest piyasa fiyat sistemi olmadan, üretim faktörlerinin değerini belirlemek mümkün değildir. Bu nedenle, merkezi planlamada doğru bilgiye ulaşmak ve karar vermek mümkün olmaz. Bu kavram, sosyalizmin pratik uygulanabilirliğini

hükümetlerin ekonomik politikalarını böyle bir varlığa devredeceğini (veya devretmesi gerektiğini), bunu yapabilseler bile, düşünmüyorum. Ayrıca, etkili ancak şüphe duyabilecekleri tedavileri almaları için insanları nasıl ikna edeceğimiz gibi sorunlar da var.

Gelişmekte olan dünyanın karşı karşıya olduğu zorluklar, hem özel hem de kamu sektörlerindeki yaygın yolsuzluk nedeniyle daha da karmaşık hale geliyor. Yolsuzluk, Bu bir kısır döngü: yoksulluğu daha da kötüleştiriyor ve yoksulluk da daha fazla yolsuzluğa yol açıyor. Yapay zekâ destekli ekonomik kalkınma planları yolsuzluk, zayıf kurumlar ve diğer birçok insani zorluklarla yüzleşmek zorunda.

Yine de iyimser olmak için önemli nedenler görüyorum. Hastalıklar ortadan kaldırıldı ve birçok ülke yoksulluktan zenginliğe geçti ve bu görevlerde yer alan kararların zekâya yüksek getiri sağladığı açık (insan kısıtlamalarına ve karmaşıklığına rağmen). Bu nedenle, yapay zekâ muhtemelen bunları şu anda yapıldığından daha iyi yapabilir.

İnsan kısıtlamalarını aşan ve yapay zekânın odaklanabileceği hedefli müdahaleler de olabilir. Ancak daha da önemlisi, denememiz gerekiyor. Hem yapay zekâ şirketleri hem de gelişmiş dünya politika yapıcıları, gelişmekte olan dünyanın dışarıda kalmamasını sağlamak için üzerlerine düşeni yapmak zorunda kalacaklar; ahlaki zorunluluk çok büyük. Bu nedenle bu bölümde iyimser argümanı savunmaya devam edeceğim, ancak her yerde başarının garanti olmadığını ve kolektif çabalarımıza bağlı olduğunu unutmayın.

---

sorgulayan temel eleştirilerden biridir.

Aşağıda, güçlü yapay zekâ geliştirildikten sonraki 5-10 yıl içinde gelişmekte olan dünyada işlerin nasıl gidebileceğine dair bazı tahminlerde bulunuyorum:

Sağlık müdahalelerinin dağıtımı. Belki de en iyimser olduğum alan, sağlık müdahalelerinin dünya çapında dağıtılmasıdır. Hastalıklar aslında yukarıdan aşağıya kampanyalarla ortadan kaldırıldı: çiçek hastalığı 1970'lerde tamamen ortadan kaldırıldı ve çocuk felci ve Gine kurdu yılda 100'den az vaka ile neredeyse ortadan kaldırıldı[11].

Matematiksel olarak gelişmiş epidemiyolojik modelleme, hastalık ortadan kaldırma kampanyalarında aktif bir rol oynamaktadır ve insanlardan daha zeki yapay zeka sistemlerinin bunu insanlardan daha iyi yapabileceği çok muhtemel görünmektedir. Dağıtım lojistiği de büyük ölçüde optimize edilebilir. GiveWell'e erken bağışçı olarak öğrendiğim bir şey, bazı sağlık yardım kuruluşlarının diğerlerinden çok daha etkili olduğudur[12]; umut, yapay zeka destekli çabaların daha da etkili olacağıdır. Ayrıca, bazı biyolojik gelişmeler dağıtım lojistiğini çok daha kolay hale getiriyor: örneğin, sıtmanın ortadan

11 Gine kurdu, kirlenmiş içme suyuyla bulaşan paraziter bir hastalıktır. Bulaşmadan 10 ila 14 ay sonra genellikle alt bacaktan yavaşça ortaya çıkar. 2025 yılında dünya genelinde yalnızca 10 vaka kaydedildi ve bu durum, çiçek hastalığından sonra ikinci insan hastalığının tamamen yok edilmesi yönünde ilerlediğini gösteriyor. Bu gelişmeler, halk sağlığı önlemleri ve erken teşhisin etkinliğini vurguluyor.

12 GiveWell, veriye dayalı ve maliyet etkinliği odaklı kurumlar arası değerlendirme yaparak bağış önerileri sunan bir nonprofit kuruluştur. Open Philanthropy ile iş birliği içinde olup, küresel sağlık ve çeşitli politika alanlarına katkı sağlayan projeler finanse eder. Bağışlar ABD ve Hollanda'da vergi indirimiyle, İngiltere'de ise Gift Aid ile desteklenir. Tüm ödeme işlemlerinde uçtan uca şifreleme kullanılır. Kuruluş, bağışların en etkili şekilde kullanılmasını sağlamayı hedefler.

kaldırılması zor olmuştur, çünkü hastalık her bulaştığında tedavi gerektirir; yalnızca bir kez uygulanması gereken bir aşı, lojistiği çok daha basitleştirir (ve sıtma için bu tür aşılar şu anda geliştirilmektedir). Daha da basit dağıtım mekanizmaları mümkün: bazı hastalıklar prensipte hayvan taşıyıcılarını hedef alarak ortadan kaldırılabilir; örneğin, hastalığı taşıma yeteneklerini engelleyen bir bakteriyle enfekte edilmiş sivrisinekleri salarak (bu sivrisinekler daha sonra diğer tüm sivrisinekleri enfekte eder) veya sivrisinekleri yok etmek için gen sürücüleri kullanarak.

Bu, milyonlarca insanı bireysel olarak tedavi etmesi gereken koordineli bir kampanya yerine, bir veya birkaç merkezi eylem gerektirir. Genel olarak, yapay zekâ destekli sağlık faydalarının iyi bir kısmının (belki %50'sinin) dünyanın en yoksul ülkelerine bile yayılması için 5-10 yılın makul bir zaman dilimi olduğunu düşünüyorum. İyi bir hedef, güçlü yapay zekânın 5-10 yıl sonra gelişmekte olan dünyanın en azından gelişmiş dünyanın bugün olduğundan önemli ölçüde daha sağlıklı olması olabilir, gelişmiş dünyanın gerisinde kalmaya devam etse bile. Bunu başarmak elbette küresel sağlık, hayırseverlik, siyasi savunuculuk ve diğer birçok alanda büyük bir çaba gerektirecektir; hem yapay zekâ geliştiricileri hem de politika yapıcılar bu konuda yardımcı olmalıdır.

Ekonomik büyüme. Gelişmekte olan dünya, sadece sağlık alanında değil, genel olarak ekonomik olarak da gelişmiş dünyaya hızla yetişebilir mi? Bunun için bazı emsaller var: 20. yüzyılın son on yıllarında, birkaç Doğu Asya ekonomisi, sürdürülebilir yıllık reel GSYİH büyüme oranlarında yaklaşık %10'luk bir artış elde ederek gelişmiş dünyaya yetişmeyi başardı. İnsan ekonomik planlamacıları, bu başarıya yol açan kararları, doğrudan tüm ekonomileri kontrol ederek değil, birkaç önemli

kaldıraç kullanarak (örneğin, ihracata dayalı büyüme odaklı bir sanayi politikası ve doğal kaynak zenginliğine güvenme cazibesine direnmek gibi) aldılar; “yapay zekalı finans bakanları ve merkez bankacıları”nın bu %10'luk başarıyı tekrarlayabileceği veya aşabileceği muhtemeldir.

Önemli bir soru, gelişmekte olan dünya hükümetlerinin bunları benimsemesini sağlarken, kendi kaderini tayin etme ilkesine saygı duymanın nasıl sağlanacağıdır; bazıları bu konuda hevesli olabilir, ancak diğerleri muhtemelen şüpheci olacaktır. İyimser taraftan bakıldığında, önceki madde işaretindeki sağlık müdahalelerinin birçoğunun organik olarak ekonomik büyümeyi artırması muhtemeldir: AIDS/sıtma/parazit solucanların ortadan kaldırılması, verimlilik üzerinde dönüştürücü bir etkiye sahip olacaktır; ayrıca bazı nörobilim müdahalelerinin (örneğin ruh halinin ve odaklanmanın iyileştirilmesi) gelişmiş ve gelişmekte olan dünyada aynı şekilde ekonomik faydaları da olacaktır. Son olarak, sağlık dışı yapay zeka destekli teknolojiler (örneğin enerji teknolojisi, ulaşım dronları, geliştirilmiş yapı malzemeleri, daha iyi lojistik ve dağıtım vb.) doğal olarak dünyaya yayılabilir; örneğin, cep telefonları bile hayırsever çabalara gerek kalmadan, piyasa mekanizmaları aracılığıyla Sahra Altı Afrika'ya hızla yayıldı. Daha olumsuz bir açıdan bakıldığında, yapay zeka ve otomasyonun birçok potansiyel faydası olsa da, özellikle henüz sanayileşmemiş ülkeler için ekonomik kalkınma açısından zorluklar da oluşturmaktadır. Bu ülkelerin artan otomasyon çağında ekonomilerini geliştirmeye devam edebilmelerini sağlamanın yollarını bulmak, ekonomistler ve politika yapıcılar için önemli bir zorluktur. Genel olarak, ideal bir senaryo—belki de hedeflenecek bir amaç—gelişmekte olan dünyada yıllık %20 GSYİH büyüme oranıdır; bunun %10'u yapay zeka destekli ekonomik

kararlardan, %10'u ise sağlık da dahil olmak üzere yapay zeka hızlandırılmış teknolojilerin doğal yayılımından kaynaklanmalıdır. Eğer bu başarılsa, Sahra altı Afrika'yı 5-10 yıl içinde Çin'in mevcut kişi başına düşen GSYİH seviyesine getirirken, gelişmekte olan dünyanın büyük bir bölümünü de mevcut ABD GSYİH'sinden daha yüksek seviyelere çıkaracaktır. Yine de, bu bir hayal senaryosu, varsayılan olarak gerçekleşen bir şey değil: hepimizin daha olası hale getirmek için birlikte çalışması gereken bir şey.

Gıda güvenliği. Daha iyi gübreler ve böcek ilaçları, daha fazla otomasyon ve daha verimli arazi kullanımı gibi mahsul teknolojisindeki gelişmeler 20. yüzyıl boyunca verimi önemli ölçüde artırarak milyonlarca insanı açlıktan kurtardı. Genetik mühendisliği şu anda birçok ürünü daha da geliştiriyor. Bunu yapmanın daha da fazla yolunu bulmak -ve tarımsal tedarik zincirlerini daha da verimli hale getirmek- bize yapay zeka destekli ikinci bir Yeşil Devrim[13] sağlayabilir ve gelişmekte olan ve gelişmiş dünya arasındaki uçurumu kapatmaya yardımcı olabilir.

İklim değişikliğinin hafifletilmesi. İklim değişikliği gelişmekte olan dünyada çok daha güçlü bir şekilde hissedilecek ve kalkınmasını engelleyecektir. Yapay zekanın, atmosferdeki karbonu uzaklaştırma ve temiz enerji teknolojilerinden, karbon yoğun endüstriyel hayvancılığa olan bağımlılığımızı azaltacak laboratuvar

13 Yeşil Devrim, 20. yüzyılın ortalarında dünya genelinde tarımsal üretimi dönüştürmüş önemli bir dönüşüm sürecidir. Yüksek verimli tohumlar, kimyasal gübreler ve modern tarım teknikleri bu süreçte yaygınlaştırılmıştır. Bu yenilikler, gıda üretimi açısından büyük ilerlemeler sağladı ve açlık riskini azaltmayı hedefledi. Ancak çevresel etkiler, su kaynaklarının aşırı kullanımı ve toprak verimliliğinin azalması gibi olumsuz yanları da mevcuttur. Yeşil Devrim, özellikle Hindistan, Meksika ve Güneydoğu Asya'da yoğun şekilde uygulanmıştır

ortamında üretilen ete kadar, iklim değişikliğini yavaşlatan veya önleyen teknolojilerde iyileşmelere yol açmasını bekleyebiliriz. Elbette, yukarıda tartışıldığı gibi, iklim değişikliği konusunda ilerlemeyi kısıtlayan tek şey teknoloji değil; bu makalede ele alınan diğer tüm konularda olduğu gibi, insan toplumsal faktörler de önemli. Ancak, yapay zeka destekli araştırmaların iklim değişikliğini hafifletmeyi çok daha az maliyetli ve derinden değiştirici[14] hale getireceğine, birçok itirazı geçersiz kılacağına ve gelişmekte olan ülkelerin daha fazla ekonomik ilerleme kaydetmesine olanak sağlayacağına dair iyi bir neden var.

Ülkeler içindeki eşitsizlik. Çoğunlukla eşitsizliği küresel bir olgu olarak ele aldım (ki bence bu onun en önemli tezahürüdür), ancak elbette eşitsizlik ülkeler içinde de mevcuttur. Gelişmiş sağlık müdahaleleri ve özellikle yaşam süresinde veya bilişsel geliştirme ilaçlarında radikal artışlarla birlikte, bu teknolojilerin "sadece zenginler için" olduğu yönünde haklı endişeler olacaktır. Özellikle gelişmiş dünyada ülke içi eşitsizlik konusunda daha iyimserim, iki nedenden dolayı. Birincisi, piyasalar gelişmiş dünyada daha iyi işliyor ve piyasalar genellikle yüksek değerli teknolojilerin maliyetini zaman içinde düşürmede başarılıdır.

İkincisi, gelişmiş ülkelerdeki siyasi kurumlar vatandaşlarına daha duyarlıdır ve evrensel erişim

14 Disruptif teknoloji, pazarları ve var olan sektörleri derinden değiştiren yeniliklerdir. Genellikle daha uygun maliyetli veya daha etkili çözümler sunarak mevcut piyasaları devirdikleri için yeni pazarlar yaratır. Bu teknolojiler, çoğunlukla düşük gelirli veya ihmal edilmiş hedef kitlelere odaklanarak başlar ve zamanla ana piyasayı ele geçirir. Örnekler arasında dijital fotoğrafçılık, bulut bilişim ve elektrikli araçlar bulunur. Disruptif teknolojiler, iş modellerini, tüketici davranışlarını ve rekabet dinamiklerini köklü şekilde etkiler.

programlarını uygulama konusunda daha büyük devlet kapasitesine sahiptirler; ve vatandaşların yaşam kalitesini bu kadar kökten iyileştiren teknolojilere erişim talep etmelerini bekliyorum. Elbette bu tür taleplerin başarılı olması önceden belirlenmiş değildir ve burada adil bir toplumu sağlamak için hep birlikte elimizden gelen her şeyi yapmalıyız. Servet eşitsizliğinde (hayat kurtaran ve yaşamı iyileştiren teknolojilere erişim eşitsizliğinin aksine) ayrı bir sorun vardır ki bu daha zor görünmektedir ve bunu 5. Bölümde ele alıyorum.

Vazgeçme sorunu. Hem gelişmiş hem de gelişmekte olan ülkelerde ortak bir endişe, insanların yapay zekâ destekli faydalardan vazgeçmesidir (aşı karşıtı hareket veya daha genel olarak Luddit hareketlerine benzer şekilde[15]). Örneğin, iyi karar verme yeteneği en düşük olan kişilerin, karar verme yeteneklerini geliştiren teknolojilerden uzak durması gibi kötü geri bildirim döngüleri ortaya çıkabilir; bu da giderek artan bir uçuruma ve hatta distopik bir alt sınıfın oluşmasına yol açabilir (bazı araştırmacılar bunun demokrasiyi baltalayacağını savunmuştur, bu konuyu bir sonraki bölümde daha ayrıntılı olarak ele alacağım). Bu, yapay zekanın olumlu ilerlemelerine ahlaki bir leke daha vuracaktır. Bu, çözülmesi zor bir problemdir çünkü insanları zorlamanın etik olarak doğru olduğunu düşünmüyorum, ancak en azından insanların bilimsel anlayışını artırmaya

15 Luddite hareketi, 1811'de Nottingham çevresinde başlayıp ardından Yorkshire, Lancashire, Derbyshire ve Leicestershire'ye yayılan, makineleri yıkan işçiler tarafından yönetilen bir isyan hareketiydi. Kumaş ve iplik sanayisindeki makine yıkımına neden olan bu olaylar 'Luddist ayaklanmaları' olarak bilinir. Hareket, 'General Ned Ludd' adlı bir figüreyle ilişkilendirilir, ancak bu kişinin gerçekliği tartışmalıdır. Kısa süreli olsa da Luddite terimi, modern çağda teknolojiye karşı direnişi simgelemek için kullanılmaya devam etmektedir.

çalışabiliriz ve belki de yapay zeka bize bu konuda yardımcı olabilir. Umut verici bir işaret, tarihsel olarak teknoloji karşıtı hareketlerin daha çok laf kalabalığı yapmasıdır: modern teknolojiye karşı çıkmak popülerdir, ancak çoğu insan sonunda onu benimser, en azından bireysel bir seçim söz konusu olduğunda. Bireyler çoğu sağlık ve tüketici teknolojisini benimseme eğilimindeyken, nükleer enerji gibi gerçekten engellenen teknolojiler genellikle kolektif siyasi kararlardır.

Genel olarak, yapay zekanın biyolojik ilerlemelerini gelişmekte olan dünyadaki insanlara hızla ulaştırma konusunda iyimserim. Yapay zekanın, benzeri görülmemiş ekonomik büyüme oranlarını mümkün kılabileceğine ve gelişmekte olan dünyanın en azından gelişmiş dünyanın şu anki durumunu aşmasına olanak sağlayabileceğine dair umudum var, ancak kesin bir inancım yok. Hem gelişmiş hem de gelişmekte olan dünyada "vazgeçme" sorunundan endişe duyuyorum, ancak bunun zamanla azalacağını ve yapay zekanın bu süreci hızlandırmaya yardımcı olabileceğini düşünüyorum. Mükemmel bir dünya olmayacak ve geride kalanlar, en azından ilk birkaç yılda, tam olarak yetişemeyecekler. Ancak bizim tarafımızdan gösterilecek güçlü çabalarla, işleri doğru yönde ve hızlı bir şekilde ilerletebiliriz. Bunu başarabilirsek, yeryüzündeki her insana borçlu olduğumuz onur ve eşitlik vaatlerinin en azından bir kısmını yerine getirebiliriz.

Barış ve Yönetişim

İlk üç bölümdeki her şeyin yolunda gittiğini varsayalım: hastalık, yoksulluk ve eşitsizlik önemli ölçüde azalır ve insan deneyiminin temel seviyesi önemli ölçüde yükselir. Bu, insan acısının tüm önemli nedenlerinin çözüldüğü anlamına gelmez. İnsanlar hâlâ birbirleri için bir tehdit oluşturuyor. Teknolojik gelişmeler ve ekonomik

kalkınmanın demokrasi ve barışa yol açtığı bir eğilim olsa da, bu çok gevşek bir eğilim ve sık sık (ve yakın zamanda) gerilemeler yaşanıyor. 20. yüzyılın başlarında insanlar savaşı geride bıraktıklarını düşünmüşlerdi; sonra iki dünya savaşı geldi. Otuz yıl önce Francis Fukuyama "Tarihin Sonu" ve liberal demokrasinin nihai zaferi hakkında yazmıştı; bu henüz gerçekleşmedi. Yirmi yıl önce ABD'li politika yapıcılar, Çin ile serbest ticaretin, Çin zenginleştikçe liberalleşmesine neden olacağına inanıyorlardı; bu hiç gerçekleşmedi ve şimdi yeniden canlanan otoriter bir blokla ikinci bir Soğuk Savaş'a doğru gidiyor gibi görünüyoruz. Ve akla yatkın teoriler, internet teknolojisinin başlangıçta inanıldığı gibi demokrasiye değil, otoriterliğe fayda sağlayabileceğini öne sürüyor (örneğin "Arap Baharı" döneminde). Güçlü yapay zekanın barış, demokrasi ve özgürlük konularıyla nasıl kesişeceğini anlamaya çalışmak önemli görünüyor.

Ne yazık ki, yapay zekanın insan sağlığını geliştireceği ve yoksulluğu hafifleteceği gibi, demokrasiyi ve barışı da yapısal olarak destekleyeceğine dair güçlü bir neden göremiyorum. İnsan çatışması düşmancadır ve yapay zeka prensipte hem "iyi adamlara" hem de "kötü adamlara" yardımcı olabilir. Hatta bazı yapısal faktörler endişe verici görünüyor: Yapay zeka, otokratların araç setindeki önemli araçlar olan propaganda ve gözetimi çok daha iyi hale getirebilir. Bu nedenle, işleri doğru yöne çevirmek bireysel aktörler olarak bize kalmış: Yapay zekanın demokrasiyi ve bireysel hakları desteklemesini istiyorsak, bu sonuç için mücadele etmemiz gerekecek. Uluslararası eşitsizlikten bile daha çok bu konuda aynı fikirdeyim: Liberal demokrasinin ve siyasi istikrarın zaferi garanti değil, belki de olası bile değil ve geçmişte olduğu gibi hepimizin büyük fedakarlık ve bağlılığını gerektirecektir.

Konunun iki bölümden oluştuğunu düşünüyorum: uluslararası çatışma ve ulusların iç yapısı. Uluslararası tarafta, güçlü yapay zekâ yaratıldığında demokrasilerin dünya sahnesinde üstünlüğe sahip olması çok önemli görünüyor. Yapay zekâ destekli otoriterlik düşünülmesi bile çok korkunç, bu nedenle demokrasilerin, hem otoriterler tarafından alt edilmekten kaçınmak hem de otoriter ülkelerde insan hakları ihlallerini önlemek için, güçlü yapay zekânın dünyaya nasıl getirileceğine dair şartları belirleyebilmesi gerekiyor.

Şu anki tahminim, bunu yapmanın en iyi yolunun bir "antente stratejisi"[16] aracılığıyla olduğu yönünde; bu stratejide demokrasilerden oluşan bir koalisyon, güçlü yapay zekâya karşı açık bir avantaj (geçici bile olsa) elde etmek için tedarik zincirini güvence altına almayı, hızla ölçeklendirmeyi ve düşmanların çip ve yarı iletken ekipman gibi kilit kaynaklara erişimini engellemeyi veya geciktirmeyi hedefliyor. Bu koalisyon bir yandan yapay zekâyı kullanarak güçlü askeri üstünlük (sopa) elde ederken, aynı zamanda güçlü yapay zekânın faydalarını (havuç) daha geniş bir ülke grubuna dağıtmayı teklif ediyor; bu da koalisyonun demokrasiyi teşvik etme stratejisini desteklemeleri karşılığında gerçekleşiyor (bu biraz "Atoms for Peace"[17]a benziyor). Koalisyon,

16 Entente stratejisi, devletlerin birbirine bağlanmasını sağlayan, hem tutukluluğu önlemek hem de korkutuculuğu korumak amacıyla geliştirilen dengeli bir yaklaşım sunar. Bu strateji, özellikle Rusya, Çin, İran ve Kuzey Kore gibi aktörlerin dış politikalarını, diplomatik ilişkilerini ve savunma alanlarını koordine etmelerinde görülmektedir. Stratejik belirsizlik, bu tür ententelerin temel özelliğidir; amaç, rakipleri belirsizliğe sokarak tepki verme zorunda bırakmaktır. Geçmişte Birinci Dünya Savaşı sırasında İttifaklar ve Koalisyonlar arasında da benzer stratejik yaklaşımlar izlenmiştir.

17 1953 yılında Başkan Eisenhower, Birleşmiş Milletler Genel Kurulu'nda yaptığı 'Atoms for Peace' konuşmasında nükleer

dünyanın giderek daha fazla kesiminin desteğini kazanmayı, en büyük rakiplerimizi izole etmeyi ve sonunda onları dünyanın geri kalanıyla aynı pazarlığı yapmanın daha iyi olacağı bir konuma getirmeyi hedefleyecektir: tüm faydaları almak ve üstün bir düşmanla savaşmamak için demokrasilerle rekabetten vazgeçmek.

Eğer tüm bunları başarabilirsek, demokrasilerin dünya sahnesinde liderlik ettiği ve otokrasiler tarafından baltalanmaktan, fethedilmekten veya sabote edilmekten kaçınacak ekonomik ve askeri güce sahip olduğu ve yapay zeka üstünlüklerini kalıcı bir avantaja dönüştürebileceği bir dünyaya sahip olacağız. Bu, iyimser bir bakış açısıyla, demokrasilerin üstün olduğu ve Fukuyama'nın hayallerinin gerçekleştiği bir "sonsuz 1991"e yol açabilir[18].

Yine de, bunu başarmak çok zor olacak ve özellikle özel yapay zeka şirketleri ile demokratik hükümetler

---

teknolojinin barışçıl amaçlar için kullanılmasını önerdi. Bu söylev, nükleer silahların yıkıcı potansiyelini vurgularken, enerji üretimi, tıp ve bilimsel araştırmada nükleer enerjinin faydalarını öne çıkardı. Bu vizyon, Uluslararası Atom Enerjisi Ajansı'nın (IAEA) kurulmasına yol açtı ve dünyada nükleer teknolojinin kontrol altında barışa hizmet etmesi yönünde önemli bir adım oldu.

18 “Eternal 1991” ifadesi, 1991 yılının dünya siyasetindeki sembolik anlamına gönderme yapıyor. 1991, Sovyetler Birliği’nin dağılmasıyla Soğuk Savaş’ın sona erdiği ve liberal demokrasilerin küresel ölçekte üstünlük kazandığı bir dönüm noktasıydı. Francis Fukuyama’nın Tarihin Sonu tezinde öne sürdüğü gibi, bu yıl liberal demokrasinin nihai zaferi ve ideolojik rekabetin bitişi olarak görülmüştü. Dolayısıyla “eternal 1991” ifadesi, iyimser bir senaryoda dünyanın sürekli olarak 1991’deki o güç dengesi halinde kalmasını, yani demokrasilerin daima üstün olduğu ve otoriter rejimlerin gerilediği bir düzenin sürmesini kast ediyor. Başka bir deyişle, bu metafor, tarihin o kritik anının kalıcılaşması ve Fukuyama’nın hayalini kurduğu demokratik düzenin “ebedileşmesi” anlamına geliyor.

arasında yakın iş birliğini ve havuç ile sopa arasındaki denge konusunda son derece akıllıca kararları gerektirecektir.

Tüm bunlar yolunda gitse bile, her ülke içinde demokrasi ve otokrasi arasındaki mücadele sorunu devam etmektedir.

Burada ne olacağını tahmin etmek elbette zor, ancak demokrasilerin en güçlü yapay zekayı kontrol ettiği küresel bir ortamda, yapay zekanın aslında her yerde demokrasiyi yapısal olarak destekleyebileceği konusunda biraz iyimserliğim var. Özellikle bu ortamda demokratik hükümetler üstün yapay zekâlarını kullanarak bilgi savaşını kazanabilirler: otokrasilerin etki ve propaganda operasyonlarına karşı koyabilirler ve hatta otokrasilerin engelleme veya izleme konusunda teknik yetenekten yoksun olduğu bir şekilde bilgi kanalları ve yapay zekâ hizmetleri sağlayarak küresel olarak özgür bir bilgi ortamı yaratabilirler. Muhtemelen propaganda yapmak gerekli değil, sadece kötü niyetli saldırılara karşı koymak ve bilgi akışının önünü açmak yeterli. Hemen olmasa da, bu tür bir eşit oyun alanı, çeşitli nedenlerle küresel yönetimi yavaş yavaş demokrasiye doğru eğme şansına sahip.

Birincisi, 1-3. Bölümlerdeki yaşam kalitesindeki artışlar, her şey eşit olduğunda, demokrasiyi teşvik etmelidir: tarihsel olarak, en azından bir ölçüde teşvik etmişlerdir. Özellikle, ruh sağlığı, refah ve eğitimdeki iyileşmelerin demokrasiyi artıracağını düşünüyorum, çünkü her üçü de otoriter liderlere destekle negatif korelasyonludur. Genel olarak insanlar, diğer ihtiyaçları karşılandığında daha fazla kendilerini ifade etmek isterler ve demokrasi, diğer şeylerin yanı sıra bir kendini ifade biçimidir. Tersine, otoriterlik korku ve kızgınlıktan beslenir.

İkincisi, otoriterler sansürleyemedikleri sürece, özgür bilginin gerçekten de otoriterliği baltalama olasılığı yüksektir. Ve sansürsüz yapay zeka, bireylere baskıcı hükümetleri baltalamak için güçlü araçlar da sağlayabilir. Baskıcı hükümetler, insanlara belirli bir tür ortak bilgiyi reddederek, onların "imparatorun çıplak olduğunu" fark etmelerini engelleyerek hayatta kalırlar. Örneğin, Sırbistan'da Milošević hükümetini devirmeye yardımcı olan Srđa Popović, otoriterleri psikolojik olarak güçlerinden mahrum bırakma, büyüyü bozma ve bir diktatöre karşı destek toplama teknikleri hakkında kapsamlı yazılar yazmıştır[19]. Herkesin cebinde, diktatörlerin engelleyemeyeceği veya sansürleyemeyeceği, insanüstü derecede etkili bir yapay zeka versiyonu (becerileri zekâya yüksek getiri sağlıyor gibi görünüyor), muhaliflerin ve reformcuların arkasında bir rüzgar yaratabilir. Tekrar söylemek gerekirse, bu uzun sürecek bir mücadele olacak, zaferin garanti olmadığı bir mücadele; ancak yapay zekayı doğru şekilde tasarlayıp inşa edersek, en azından her yerde özgürlük savunucularının avantajlı olduğu bir mücadele olabilir.

Sinirbilim ve biyolojide olduğu gibi, işlerin nasıl "normalden daha iyi" olabileceğini de sorabiliriz—sadece otokrasiden nasıl kaçınılacağını değil, demokrasileri bugün olduklarından nasıl daha iyi hale getirebileceğimizi de. Demokrasilerde bile, adaletsizlikler her zaman yaşanır. Hukuk devleti toplumları, vatandaşlarına herkesin kanun önünde eşit olacağına ve herkesin temel insan haklarına

19 Srđa Popović, Sırbistan'ın önde gelen avukat ve sivil toplum figürü, 29 Ekim 2013'te Belgrad'da 76 yaşında vefat etti. 2011 yılında 'Global Thinkers' listesinde yer alarak Arap Baharı'na ilham vermesi nedeniyle tanındı. 2012'de Nobel Barış Ödülü adaylığı aldı. Halkın barışçıl direniş stratejileri konusunda uzmanlaşmış, Sırbistan'da insan hakları ve demokratik dönüşüm için büyük katkılar sağladı.

sahip olacağına dair bir söz verir, ancak açıkçası insanlar pratikte her zaman bu hakları almazlar. Bu vaadin kısmen bile olsa yerine getirilmesi gurur duyulacak bir şey, ancak yapay zeka daha iyisini yapmamıza yardımcı olabilir mi?

Örneğin, yapay zeka kararları ve süreçleri daha tarafsız hale getirerek hukuk ve yargı sistemimizi iyileştirebilir mi? Bugün insanlar çoğunlukla hukuk veya yargı bağlamlarında yapay zeka sistemlerinin ayrımcılığa neden olacağından endişe ediyorlar ve bu endişeler önemli ve bunlara karşı savunma yapılması gerekiyor. Aynı zamanda, demokrasinin canlılığı, sadece risklere yanıt vermekle kalmayıp, demokratik kurumları iyileştirmek için yeni teknolojilerden yararlanmaya bağlıdır. Yapay zekanın gerçekten olgun ve başarılı bir uygulaması, önyargıyı azaltma ve herkes için daha adil olma potansiyeline sahiptir.

Yüzyıllardır, hukuk sistemleri, yasanın tarafsız olmayı hedeflediği, ancak doğası gereği öznel olduğu ve bu nedenle önyargılı insanlar tarafından yorumlanması gerektiği ikilemiyle karşı karşıya kalmıştır. Yasayı tamamen mekanik hale getirmeye çalışmak işe yaramadı çünkü gerçek dünya karmaşık ve her zaman matematiksel formüllerle yakalanamaz. Bunun yerine hukuk sistemleri, "zalim ve alışılmadık ceza" veya "tamamen kurtarıcı sosyal önemi olmayan" gibi son derece belirsiz kriterlere dayanır ve insanlar bunları yorumlar—ve genellikle önyargı, kayırmacılık veya keyfilik sergileyen bir şekilde yaparlar. Kripto paralardaki "akıllı sözleşmeler" hukuku devrimleştirmemiştir çünkü sıradan kod, ilgi çekici birçok şeyi yargılayacak kadar akıllı değildir. Ancak yapay zeka bunun için yeterince akıllı olabilir: Tekrarlanabilir ve mekanik bir şekilde geniş, belirsiz kararlar verebilen ilk teknolojidir.

Hakimleri kelimenin tam anlamıyla yapay zeka sistemleriyle değiştirmeyi önermiyorum, ancak tarafsızlığın, karmaşık, gerçek dünya durumlarını anlama ve işleme yeteneğiyle birleşimi, hukuk ve adalete ciddi olumlu uygulamalar getirmesi gerektiğini düşündürüyor. En azından, bu tür sistemler, karar vermede insanlara yardımcı olarak çalışabilir. Bu tür herhangi bir sistemde şeffaflık önemli olacaktır ve olgun bir yapay zeka bilimi bunu sağlayabilir: bu tür sistemlerin eğitim süreci kapsamlı bir şekilde incelenebilir ve gelişmiş yorumlanabilirlik teknikleri, nihai modelin içine bakmak ve onu gizli önyargılar açısından değerlendirmek için kullanılabilir; bu, insanlarda mümkün olmayan bir şekilde yapılabilir. Bu tür yapay zeka araçları, temel hakların ihlallerini izlemek için de kullanılabilir ve anayasaları daha kendi kendini uygulayan hale getirebilir.

Benzer şekilde, yapay zeka hem görüşleri bir araya getirmek hem de vatandaşlar arasında fikir birliği sağlamak, çatışmaları çözmek, ortak zemin bulmak ve uzlaşma aramak için kullanılabilir.

Bu yöndeki bazı erken fikirler, Anthropic ile yapılan iş birlikleri de dahil olmak üzere, hesaplamalı demokrasi projesi tarafından hayata geçirildi. Daha bilgili ve düşünceli bir vatandaşlık, demokratik kurumları açıkça güçlendirecektir. Ayrıca, yapay zekanın, prensipte herkese açık olan ancak pratikte genellikle ciddi şekilde eksik olan ve bazı yerlerde diğerlerinden daha kötü olan sağlık hizmetleri veya sosyal hizmetler gibi devlet hizmetlerinin sağlanmasına yardımcı olmak için kullanılması için açık bir fırsat vardır.

Bu, sağlık hizmetlerini, trafik tescil dairesini, vergileri, sosyal güvenliği, bina kodu uygulamasını vb. İçerir. Çok düşünceli ve bilgili bir yapay zekânın görevinin size hükümet tarafından yasal olarak hak ettiğiniz her şeyi

anlayabileceğiniz bir şekilde vermek ve aynı zamanda genellikle kafa karıştırıcı olan hükümet kurallarına uymanıza yardımcı olmak olması büyük bir olay olurdu. Devlet kapasitesinin artırılması hem yasa önünde eşitlik vaadini yerine getirmeye yardımcı olur hem de demokratik yönetime saygıyı güçlendirir. Kötü uygulanan hizmetler şu anda hükümete karşı duyulan şüpheciliğin önemli bir nedenidir.

Bunların hepsi biraz belirsiz fikirler ve bu bölümün başında da söylediğim gibi, bunların uygulanabilirliğine biyoloji, nörobilim ve yoksulluğun azaltılmasındaki ilerlemeler kadar güvenmiyorum. Gerçekçi olmayacak şekilde ütopik olabilirler. Ama önemli olan iddialı bir vizyona sahip olmak, büyük hayaller kurmaya ve yeni şeyler denemeye istekli olmaktır. Vizyon Yapay zekanın özgürlüğün, bireysel hakların ve hukuk önünde eşitliğin garantörü olarak rolü, uğruna mücadele edilmeyecek kadar güçlü bir vizyondur. 21. yüzyılda, yapay zeka destekli bir siyasi sistem, hem bireysel özgürlüğün daha güçlü bir koruyucusu hem de liberal demokrasinin tüm dünyanın benimsemek istediği yönetim biçimi haline gelmesine yardımcı olan bir umut ışığı olabilir.

İş ve Anlam

Önceki dört bölümdeki her şey yolunda gitse bile—sadece hastalığı, yoksulluğu ve eşitsizliği hafifletmekle kalmayıp, liberal demokrasi baskın yönetim biçimi haline gelse ve mevcut liberal demokrasiler kendilerinin daha iyi versiyonları haline gelse bile—en az bir önemli soru hâlâ cevapsız kalıyor. "Böylesine teknolojik olarak gelişmiş, adil ve dürüst bir dünyada yaşıyor olmamız harika," diye itiraz edebilir biri, "ama yapay zekâ her şeyi yaparken, insanların nasıl bir anlamı olacak? Dahası, ekonomik olarak nasıl hayatta kalacaklar?".

Bence bu soru diğerlerinden daha zor. Bununla, diğer sorulara göre bu konuda mutlaka daha kötümser olduğum anlamına gelmiyor (zorluklar görsem de). Yani daha belirsiz ve önceden tahmin edilmesi daha zor, çünkü toplumun nasıl örgütlendiğine dair makroskopik sorularla ilgili ve bu sorular genellikle zaman içinde ve merkezi olmayan bir şekilde kendiliğinden çözülüyor. Örneğin, tarihsel avcı-toplayıcı toplumlar, hayatın avcılık ve çeşitli avcılıkla ilgili dini ritüeller olmadan anlamsız olduğunu düşünmüş olabilirler ve iyi beslenmiş teknolojik toplumumuzun amaçsız olduğunu düşünmüş olabilirler. Ayrıca, ekonomimizin herkesi nasıl geçindirebileceğini veya mekanize bir toplumda insanların hangi işlevi faydalı bir şekilde yerine getirebileceğini anlamamış olabilirler.
Yine de, bu bölümün kısalığının bu konuları ciddiye almadığımın bir işareti olarak algılanmaması gerektiğini, aksine net cevapların eksikliğinin bir işareti olduğunu unutmadan, en azından birkaç kelime söylemeye değer.

Anlam sorusuna gelince, bir yapay zekanın daha iyi yapabileceği için üstlendiğiniz görevlerin anlamsız olduğuna inanmanın büyük olasılıkla bir hata olduğunu düşünüyorum. Çoğu insan hiçbir şeyde dünyanın en iyisi değildir ve bu onları pek rahatsız etmiyor gibi görünüyor. Elbette bugün hala karşılaştırmalı avantaj yoluyla katkıda bulunabilirler ve ürettikleri ekonomik değerden anlam çıkarabilirler, ancak insanlar ekonomik değer üretmeyen faaliyetlerden de büyük keyif alırlar. Ben de bolca zamanımı video oyunları oynayarak, yüzerek, dışarıda yürüyerek ve arkadaşlarımla konuşarak geçiriyorum; bunların hepsi sıfır ekonomik değer üretiyor. Bir günümü bir video oyununda daha iyi olmak veya dağa bisikletle daha hızlı tırmanmak için harcayabilirim ve bir yerlerde birilerinin bu konularda benden çok daha iyi olması benim için gerçekten önemli değil. Her durumda, anlamın

çoğunlukla insan ilişkilerinden ve bağlantılarından geldiğini düşünüyorum, ekonomik emekten değil. İnsanlar bir başarı duygusu, hatta bir rekabet duygusu isterler ve yapay zekâ sonrası bir dünyada, tıpkı insanların bugün araştırma projelerine girişirken, Hollywood oyuncusu olmaya çalışırken veya şirket kurarken yaptıkları gibi, karmaşık bir stratejiyle çok zor bir görevi yıllarca denemek tamamen mümkün olacaktır. (a) Bir yerde bir yapay zekanın prensipte bu görevi daha iyi yapabileceği gerçeği ve (b) bu görevin artık küresel ekonominin ekonomik olarak ödüllendirilen bir unsuru olmaması gerçeği bana pek önemli görünmüyor.

Ekonomik kısım aslında bana anlam kısmından daha zor geliyor. Bu bölümde "ekonomik" derken, insanların çoğunun veya tamamının yeterince gelişmiş bir yapay zeka odaklı ekonomiye anlamlı bir şekilde katkıda bulunamayabileceği olası sorunu kastediyorum. Bu, özellikle 3. Bölümde ele aldığım yeni teknolojilere erişimdeki eşitsizlik gibi ayrı bir eşitsizlik sorunundan daha makro bir sorundur.

Öncelikle, kısa vadede karşılaştırmalı avantajın insanları alakalı tutmaya devam edeceği ve aslında verimliliklerini artıracağı ve hatta bazı yönlerden insanlar arasında oyun alanını eşitleyebileceği argümanlarına katılıyorum. Yapay zekâ belirli bir işin yalnızca %90'ında daha iyi olduğu sürece, geri kalan %10'luk kısım insanların yüksek oranda kaldıraçlı hale gelmesine, ücretlerin artmasına ve aslında yapay zekânın iyi olduğu şeyleri tamamlayan ve güçlendiren bir sürü yeni insan işi yaratılmasına neden olacaktır; böylece “%10” neredeyse herkesi istihdam etmeye devam edecek şekilde genişleyecektir. Aslında, yapay zekâ her şeyi insanlardan %100 daha iyi yapabilse bile, bazı görevlerde verimsiz veya pahalı kalırsa veya insanlara ve yapay zekâya yönelik

kaynak girdileri anlamlı derecede farklıysa, o zaman karşılaştırmalı avantaj mantığı geçerli olmaya devam eder. İnsanların önemli bir süre boyunca göreceli (hatta mutlak) bir avantajı koruyabileceği bir alan fiziksel dünyadır. Bu nedenle, insan ekonomisinin, "veri merkezindeki dâhiler ülkesine" ulaştığımız noktadan biraz sonra bile anlamlı olmaya devam edebileceğini düşünüyorum.

Ancak, uzun vadede yapay zekânın o kadar yaygın olarak etkili ve o kadar ucuz hale geleceğini düşünüyorum ki, bu durum artık geçerli olmayacak. O noktada mevcut ekonomik düzen artık mantıklı olmayacak ve ekonominin nasıl organize edilmesi gerektiği konusunda daha geniş bir toplumsal tartışmaya ihtiyaç duyulacak.

Bu kulağa çılgınca gelse de, gerçek şu ki, uygarlık geçmişte büyük ekonomik değişimleri başarıyla yönetti: avcılık-toplayıcılıktan tarıma, tarımdan feodalizme ve feodalizmden endüstriyele. Sanırım yeni ve daha tuhaf bir şeye ihtiyaç duyulacak ve bugün kimsenin iyi bir şekilde tasavvur edemediği bir şey bu. Herkes için büyük bir evrensel temel gelir kadar basit olabilir, ancak bunun çözümün sadece küçük bir parçası olacağını düşünüyorum.

Bu, yapay zeka sistemlerinin kapitalist bir ekonomisi olabilir; bu sistemler daha sonra yapay zeka sistemlerinin insanlarda ödüllendirmeyi mantıklı bulduğu (nihayetinde insan değerlerinden türetilen bir yargıya dayalı olarak) ikincil bir ekonomiye göre insanlara kaynak (çok büyük miktarlarda, çünkü genel ekonomik pasta çok büyük olacak) dağıtır. Belki de ekonomi Whuffie puanları üzerinden işliyor[20]. Ya da belki de insanlar, alışılagelmiş

20 “Whuffie” kavramı, Cory Doctorow’un 2003 tarihli bilimkurgu romanı Down and Out in the Magic Kingdom’da ortaya çıkan bir fikir. Burada Whuffie, geleceğin toplumunda insanların ekonomik değerini değil, itibarlarını ve sosyal katkılarını ölçen bir tür

ekonomik modellerin öngörmediği bir şekilde, sonuçta ekonomik olarak değerli olmaya devam edecekler.

Bu çözümlerin hepsinin çok sayıda olası sorunu var ve bunların mantıklı olup olmayacağını çok fazla yineleme ve deneme yapmadan bilmek mümkün değil. Ve diğer bazı zorluklarda olduğu gibi, burada da iyi bir sonuç elde etmek için mücadele etmemiz gerekecek: sömürücü veya distopik yönler de açıkça mümkün ve önlenmeleri gerekiyor. Bu sorular hakkında çok daha fazla şey yazılabilir ve bunu daha sonraki bir zamanda yapmayı umuyorum.

Değerlendirme

Yukarıdaki çeşitli konular aracılığıyla, yapay zeka ile her şey yolunda giderse hem makul hem de bugünkü dünyadan çok daha iyi bir dünya vizyonu ortaya koymaya çalıştım. Bu dünyanın gerçekçi olup olmadığını bilmiyorum ve gerçek olsa bile, çok büyük bir çaba ve

---

"sosyal kredi" olarak işliyor. Whuffie'nin Özellikleri. Reputasyon temelli: Para yerine, insanların toplum için ne kadar değerli işler yaptığına göre puan kazanması. Toplumsal katkı: Faydalı, yaratıcı veya topluma pozitif katkı sağlayan davranışlar Whuffie puanını artırıyor. Negatif davranışlar: Antisosyal veya zararlı davranışlar puanı düşürüyor. Demokratik ölçüm: Herkesin başkalarının Whuffie'sini artırma veya azaltma hakkı var. Sonraki Kullanımlar Tara Hunt'un The Whuffie Factor kitabı (2009): Doctorow'un fikrini alıp iş dünyası ve sosyal medya bağlamında yorumladı. Burada Whuffie, şirketlerin ve bireylerin sosyal sermayesini, yani güven, itibar ve toplulukla kurduğu bağları ölçen bir metafor olarak kullanılıyor. Sosyal değer ölçütü: Modern dünyada "sosyal currency" (sosyal para birimi) kavramıyla bağlantılı; takipçi sayısı, etkileşim, güvenilirlik gibi unsurlar Whuffie'nin güncel karşılıkları olarak düşünülebilir.

Whuffie gerçek bir banka puanı ya da finansal sistem değil; itibarın ve sosyal değer üretiminin ölçülmesi için kullanılan kurgusal bir metafor. Bugün sosyal medya etkileşimleri, influencer gücü veya marka itibarı gibi kavramlar, Whuffie'nin modern dünyadaki yankıları olarak görülebilir.

birçok cesur ve özverili insanın mücadelesi olmadan başarılamayacak. Herkes (yapay zeka şirketleri de dahil!) hem riskleri önlemek hem de faydaları tam olarak gerçekleştirmek için kendi payına düşeni yapmalı.

Ama uğruna savaşmaya değer bir dünya. Eğer tüm bunlar gerçekten 5 ila 10 yıl içinde gerçekleşirse—çoğu hastalığın yenilmesi, biyolojik ve bilişsel özgürlüğün artması, milyarlarca insanın yoksulluktan kurtarılarak yeni teknolojilerden pay alması, liberal demokrasinin ve insan haklarının yeniden doğuşu—bunu izleyen herkesin bunun kendileri üzerindeki etkisine şaşıracağını düşünüyorum. Yeni teknolojilerin hepsinden kişisel olarak faydalanma deneyimini kastetmiyorum, ki bu da elbette muhteşem olacak. Kastettiğim, uzun zamandır sahip olduğumuz ideallerin bir anda gözlerimizin önünde somutlaşması deneyimi. Bence birçok kişi bundan dolayı gözyaşlarına boğulacak.

Bu denemeyi yazarken ilginç bir gerilim fark ettim. Bir anlamda burada ortaya konan vizyon son derece radikal: neredeyse hiç kimsenin önümüzdeki on yılda olmasını beklemediği bir şey ve muhtemelen birçok kişiye absürt bir fantezi gibi gelecek. Bazıları bunu arzu edilebilir bile bulmayabilir; herkesin katılmayacağı değerleri ve siyasi tercihleri içeriyor.

Ama aynı zamanda, sanki iyi bir dünya tasavvur etmeye yönelik birçok farklı girişim kaçınılmaz olarak kabaca buraya götürüyormuş gibi, bunda göz kamaştırıcı derecede açık bir şey—aşırı belirlenmiş bir şey—var.

Iain M. Banks'in "Oyunların Oyuncusu" adlı eserinde, baş kahraman (burada açıkladığım ilkelere benzer prensiplere dayanan Kültür adlı bir topluluğun üyesi), liderliğin karmaşık bir savaş oyununda rekabetle belirlendiği baskıcı, militarist bir imparatorluğa seyahat

eder[21]. Ancak oyun, bir oyuncunun stratejisinin kendi siyasi ve felsefi bakış açısını yansıtması için yeterince karmaşıktır. Kahraman, oyunda imparatoru yenmeyi başarır ve değerlerinin (Kültür'ün değerleri) acımasız rekabete ve en güçlü olanın hayatta kalmasına dayalı bir toplum tarafından tasarlanmış bir oyunda bile kazanma stratejisini temsil ettiğini gösterir.

Scott Alexander'ın tanınmış bir yazısı da aynı tezi savunuyor[22]: rekabet kendi kendini baltalar ve şefkat ve işbirliğine dayalı bir topluma yol açma eğilimindedir. “Ahlaki evrenin yayı” da benzer bir kavramdır. Bence

21 Ünlü İskoç yazar Iain M. Banks’in 1988’de yayımladığı The Player of Games adlı bilimkurgu romanına atıfta bulunuyor. Romanın bağlamı: Bu eser, Banks’in “Culture” serisinin ikinci kitabıdır. “Culture” adı verilen toplum, ileri teknolojiye sahip, post-kıtlık bir uygarlık olarak tasvir edilir; bireysel özgürlük, bolluk ve yapay zekâların rehberliği üzerine kuruludur. Konu: Başkahraman Jernau Morat Gurgeh, oyunlarda ustalaşmış bir Culture vatandaşıdır. O, Azad adlı baskıcı ve militarist bir imparatorluğa gönderilir. Bu imparatorlukta siyasi ve toplumsal düzen, karmaşık bir strateji oyunu olan “Azad” üzerinden belirlenir; liderlik ve iktidar bu oyundaki başarıya bağlıdır. Cümlenin anlamı: Dolayısıyla “In Iain M. Banks’ The Player of Games” ifadesi, “Iain M. Banks’in The Player of Games adlı romanında” demektir. Yani cümle, örnek olarak bu romandaki kurgusal durumu kullanıyor: özgürlükçü bir toplumdan gelen bir bireyin, baskıcı bir rejimde oyun yoluyla iktidarın belirlendiği düzenle karşılaşmasını anlatıyor. Bu roman, hem oyun metaforu üzerinden güç ve iktidar ilişkilerini sorgulamasıyla hem de Culture evreninin değerlerini otoriter sistemlerle karşılaştırmasıyla öne çıkar.

22 Scott Alexander, önceki adıyla Slate Star Codex (SSC), bilim, tıp (özellikle psikiyatri), felsefe, siyaset ve gelecekçilik konularında yazılar yazan pseudonim bir yazardır. Aynı zamanda bir psikiyatristtir. Blogu, teknoloji, psikoloji ve toplumsal politikalar gibi alanlarda derin analizler sunar. 2027'deki entelektüel patlama öncesi ayları inceleyen Daniel Kokotajlo ile iş birliği de yapmaktadır. Bu çalışmalar, yapay zekânın etkileri ve toplum üzerine düşüncelerini genişletmektedir.

Kültürün değerleri, açık ahlaki güce sahip ve herkesi aynı tarafa çekme eğiliminde olan milyonlarca küçük kararın toplamı olduğu için kazanan bir stratejidir.

Temel insan sezgileri olan adalet, işbirliği, merak ve özerklik ile tartışmak zordur ve daha yıkıcı dürtülerimizin aksine birikimlidirler. Çocukların hastalıktan ölmemesi gerektiğini savunmak kolaydır ve buradan herkesin çocuklarının bu hakkı eşit olarak hak ettiğini savunmak da kolaydır. Buradan, hepimizin bir araya gelip bu sonucu elde etmek için zekamızı kullanmamız gerektiğini savunmak da zor değildir. İnsanların gereksiz yere saldırmaları veya başkalarına zarar vermeleri nedeniyle cezalandırılmaları gerektiği konusunda çok az kişi hemfikir değil ve buradan cezaların tutarlı ve sistematik olması gerektiği fikrine geçmek de çok zor değil. Benzer şekilde, insanların kendi hayatları ve seçimleri üzerinde özerkliğe ve sorumluluğa sahip olmaları da sezgiseldir. Bu basit sezgiler, mantıksal sonuçlarına götürüldüğünde, sonunda hukukun üstünlüğüne, demokrasiye ve Aydınlanma değerlerine yol açar. Kaçınılmaz olmasa da, en azından istatistiksel bir eğilim olarak, insanlık zaten bu yöne doğru ilerliyordu. Yapay zeka, bizi oraya daha hızlı ulaştırmak, mantığı daha keskin ve hedefi daha net hale getirmek için bir fırsat sunuyor.

Yine de, bu olağanüstü bir güzellik. Bunu gerçeğe dönüştürmede küçük bir rol oynama fırsatımız var. Bu denemenin taslaklarını inceledikleri için Kevin Esvelt, Parag Mallick, Stuart Ritchie, Matt Yglesias, Erik Brynjolfsson, Jim McClave, Allan Dafoe ve Anthropic'teki birçok kişiye teşekkürler.

Yolun kazananlarına.